Uncertainty Propagation in Operational Flash Flood Forecasting Systems

---

Dissertation
zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich

von
Katharina Liechti
von
Landiswil BE

Promotionskomitee
Prof. Dr. Jan Seibert (Vorsitz)
Prof. Dr. Wilfried Haeberli
Dr. Massimiliano Zappa (Leitung der Dissertation)

Zürich, 2013

Contents

ABSTRACT

Mountainous catchments prone to flash floods are typically relatively small in size, have shallow soils and steep slopes. Thus, they respond quickly to heavy rainfall, and consequently the time left for warning is short. Therefore, flash flood forecasts for such catchments are needed. The IMPRINTS project aimed to improve preparedness and risk management for flash flood and debris flow events, and within this framework this dissertation aims to make advances in flash flood forecasting.

Many catchments prone to flash flood are poorly monitored or ungauged. For such catchments it is particularly difficult to make early warnings, because of the large uncertainty of the main triggering variable, the precipitation, and the absence of valuable time series of observed discharge to calibrate a hydrological model.

The uncertainties in the input data of flash flood forecasting systems accumulate with the uncertainties of the process description and parameterisation of the hydrological model and propagate through the forecasting process and are an integral part of the resulting forecast.

In this dissertation several approaches to account for these uncertainties are investigated in five test catchments in the Lago Maggiore region, in Southern Switzerland. Particular emphasis is given to the application of weather radar data, as this is regarded as a promising tool to provide forecasts also in poorly gauged and small catchments.

Indeed, it is shown that for ungauged catchments forecast chains driven by quantitative precipitation estimates (QPE) from weather radar can provide better estimates of discharge then forecasts driven by interpolated rain-gauge data, when the uncertainties in the weather radar data are accounted for by applying an ensemble of weather radar QPE.

It is also shown that uncertainties originating from the parameterisation of the hydrological model can be accounted for with an ensemble of different parameter sets, which then result in more skilful forecasts. Furthermore it is suggested to use probabilistic nowcasts, like radar ensembles, to produce ensembles of initial conditions for a subsequent initialisation of forecasts driven by deterministic high-resolution numerical weather prediction (NWP) forecasts. With such a flash flood forecast chain longer lead times are reached compared to pure nowcast products and higher forecast skill is reached compared to forecasts driven only by deterministic high resolution NWP forecasts.

ZUSAMMENFASSUNG

Für Sturzfluten anfällige Gebirgseinzugsgebiete sind meist durch eine kleine Fläche, flachgründige Böden und steile Hänge gekennzeichnet. Sie reagieren daher schnell auf Starkniederschläge, was das Zeitfenster für Warnungen sehr begrenzt. Daraus ergibt sich in solchen Einzugsgebieten die Notwendigkeit für Sturzflutvorhersagen. Das Projekt IMPRINTS hatte zum Ziel die Frühwarnung und das Risikomanagement für Sturzfluten und Murgänge zu verbessern. Im Rahmen dieses Projekts, sind Fortschritte im Bereich der Vorhersage von Sturzfluten das Ziel der vorliegenden Dissertation.

Viele für Sturzfluten anfällige Gebirgseinzugsgebiete werden nur schlecht oder gar nicht von Messnetzen abgedeckt. Die Unsicherheit in den Annahmen über den Niederschlag und das Fehlen wertvoller Zeitreihen von Abflussmessungen für die Kalibrierung von hydrologischen Modellen machen Frühwarnungen für solche Gebiete besonders schwierig.

Die Unsicherheiten in den Eingangsdaten von Hochwasservorhersagesystemen summieren sich auf mit den Unsicherheiten bezüglich der Prozessbeschreibung und Parametrisierung des hydrologischen Modells. Diese Unsicherheiten pflanzen sich durch den ganzen Vorhersageprozess fort und sind daher ein integraler Bestandteil der resultierenden Vorhersage.

In der vorliegenden Dissertation werden verschiedene Ansätze zur Berücksichtigung der beschriebenen Unsicherheiten geprüft. Dazu werden Studien in fünf Teileinzugsgebieten des Lago Maggiore, in der Südschweiz durchgeführt. Ein Schwerpunkt der Studien liegt auf der Anwendung von Wetterradardaten, da diese als ein vielversprechendes Mittel für wenig beobachtete, kleine Einzugsgebiete angesehen werden.

Tatsächlich konnte gezeigt werden, dass Abflussvorhersagen basierend auf Wetterradardaten bessere Resultate liefern als Vorhersagen basierend auf interpolierten Pluviometerdaten, sofern die Unsicherheiten in den Radardaten in Form eines Ensembles berücksichtigt werden.

Den Unsicherheiten im Zusammenhang mit der Modellparametrisierung kann durch die Anwendung eines Ensembles verschiedener Parametersätze Rechnung getragen werden. Dadurch kann die Güte der Vorhersage gesteigert werden. Ausserdem wird empfohlen probabilistische Echtzeitvorhersagen des Abflusses zu verwenden, um ein Ensemble von Anfangsbedingungen zu ermitteln. Diese werden dann für eine darauffolgende Abflussvorhersage verwendet, welche mit Daten von deterministischen hoch aufgelösten numerischen Wettermodellen angetrieben wird. Mit einer solchen Vorhersagekette für Sturzfluten können längere Vorlaufzeiten erreicht werden, als mit reinen Echtzeitprodukten. Auch werden höhere Güten als mit rein deterministischen Vorhersagen erreicht.

Uncertainty Propagation in Operational Flash Flood Forecasting Systems

This dissertation consists of four papers and an introductory part that summarises the four papers. Additionally an appendix contains further work accomplished during the preparation of this dissertation.

LIST OF PAPERS

I. Rossa A., Liechti K., Zappa M., Bruen M., Germann U., Haase G., Keil C., Krahe P., 2011. The COST 731 Action: A review on uncertainty propagation in advanced hydro-meteorological forecast systems. *Atmospheric Research* 100: 150-167.

II. Liechti K., Fundel F., Germann U., Zappa M., 2012. Flood nowcasting in the Southern Swiss Alps using radar ensemble. In *Weather Radar and Hydrology*, Moore R.J., Cole S.J., Illingworth A.J. eds). vol. 351, Red Book: Exeter, UK.

III. Liechti K., Zappa M., Fundel F., Germann U., 2013. Probabilistic evaluation of ensemble discharge nowcasts in two nested Alpine basins prone to flash floods. *Hydrological Processes* 27(1): 5-17.

IV. Liechti K., Panziera L., Germann U., Zappa M., 2013. Flash-flood early warning using weather radar data: from nowcasting to forecasting. *Hydrology and Earth System Sciences Discussion* 10: 1289-1331.

AI. Liechti K., Turowski J.M., Zappa M. Five perspectives on ranking two severe flash flood events in Switzerland. *Manuscript.*

AII. Zappa M., Jaun S., Badoux A., Schwanbeck J., Addor N., Liechti K., Roeser I., Walser A., Viviroli D., Vogt S., Gerber M., Trösch J., Weingartner R., Oplatka M., Bezzola G.R., Rhyner J., 2010. IFKIS-Hydro Sihl: Ein operationelles Hochwasservorhersagesystem für die Stadt Zürich und das Sihltal. *Wasser Energie Luft* 102(3): 238-248.

In paper I, I was responsible for the literature review as well as the main part of writing, excluding the sections specifically concerning the COST 731 Action and the section concerning radar data assimilation. In papers II, III, IV and AI, I was responsible for computations, analysis and the main part of writing. In papers IV and AI, I was also co-responsible for the experimental set up. In paper AII I was involved in parts of the analysis.

INTRODUCTION

Water belongs to the essentials of life and yet it can also be very destructive. An example for that are flash floods. When a catchment reacts very fast to an intensive rainfall event in form of a flash flood, nature demonstrates the destructive force of water. These forces of nature are both threatening and fascinating.

To keep the consequences of such events within limits, forecasts are needed. Because when an intensive rainfall event occurs over a flash flood prone catchment it is generally too late to take action or the time to do so is very limited. Mitigation action and preparation prior to the event require forecasts with adequate lead time to warn and to take action.

This is not an easy task. Flash floods are extreme events, that only happen infrequently or once every couple of years at a certain location. If a flash flood occurs, however, it often leads to severe damage. In years without large scale flooding, the highest amount of damages along Swiss rivers is caused by flash floods during the thunderstorm season in summer (Hilker *et al.*, 2009). So it is obvious that forecasts for flash floods are desirable and needed. But here lies one of the difficulties in forecasting. The forecast is most needed for events that are least foreseeable. Exactly because flash floods are extreme events, there is only little data available to test forecasting systems and to learn from. So even though flash floods are among the most serious natural hazards, which frequently cause severe damage to the environment and infrastructure, as well as loss of life (Borga et al., 2011; Chiang et al., 2007; Hapuarachchi et al., 2011), they are still scarcely documented and poorly understood natural phenomena (Gaume et al., 2009; Rusjan et al., 2009).

Flash floods are characterized by a strong response to intense rainfall, such as thunderstorms, leading to rapidly rising and falling hydrographs, resulting in limited opportunity for warnings to be prepared and issued (Collier, 2007; Perry, 2000; USGS, 2008). Mountainous catchments are particularly prone to flash floods as their topography favours heavy convective precipitation events (Panziera and Germann, 2010; Smith, 1979) and due to the shallow soils and steep slopes in such catchments the response time is generally short (Barredo, 2007). Those flashy catchments are usually small in size and often poorly gauged or ungauged, which makes it very difficult to obtain reliable estimates of precipitation (Hapuarachchi *et al.*, 2011; Werner and Cranston, 2009). Therefore, one of the main difficulties in forecasting flash floods is to obtain information about the spatial and temporal distribution, the duration and the intensity of rainfall.

There are several ways to obtain this information about precipitation, which is needed to force a rainfall-runoff model. The oldest and most established way is to measure precipitation with rain-gauges (Sevruk, 1996). A network of rain-gauges allows to generate a surface of precipitation estimates by interpolating between the rain-gauges (Thornton et al., 1997). Precipitation estimates from weather radar are particularly useful to detect convective precipitation as weather radar information is available at very high spatial and temporal resolution (Germann *et al.*, 2006). To obtain information about future precipitation, numerical weather predictions (NWP) can be used, which give an estimate of the expected precipitation for the next few days (Montani et al., 2011). However, all these methods for precipitation estimation have their limits and inherent sources of uncertainties. In hydrological modelling, these uncertainties propagate through the hydrological model and affect the resulting estimates of discharge (Zappa *et al.*, 2011).

Uncertainty in hydrological modelling

The observed or forecasted meteorological input is one of the main sources of uncertainty in flash flood early warning systems, but by far not the only one. Other sources of uncertainty lie in the definition of the initial conditions of the hydrological model and the model structure and parameterization. In the following these sources of uncertainties are described.

Sources of uncertainty

Precipitation measurements from rain-gauges are usually relatively accurate but only cover small areas of some square decimetres (Michelson, 2004; Sevruk, 1996). If the data from several rain-gauges or a whole rain-gauge network are interpolated over space to form surfaces of precipitation estimates, this means that measurements representative for only a few square meters are used to estimate the precipitation for tens or hundreds of square kilometres (Schiemann et al., 2010; Tobin et al., 2011; Velasco-Forero et al., 2009). Therefore, rainfall data from rain-gauges often capture the spatial distribution and intensity of a rainfall event insufficiently for flash flood early warning (Griffiths et al., 2009; Rusjan et al., 2009). Furthermore, precipitation estimates from rain-gauges are available in real-time at the best, and thus the lead time of a discharge forecast can only be as long as the response time of the catchment. The latter also applies for weather radar quantitative precipitation estimates (QPE). But due to their good resolution in space and time (1 km in space and 5 minutes in time in Switzerland), radar QPE are seen as a promising tool for a large variety of hydrological problems and especially for risk management in catchments with fast response times like mountainous or urban catchments (Borga *et al.*, 2007; Delrieu *et al.*, 2009; Krajewski and Smith, 2002; Morin *et al.*, 2009). But radar rainfall estimates are affected by many sources of error, which have to be dealt with. Amongst others, these are beam shielding, attenuation by water on the radome, attenuation in heavy rain and hail, overestimation in hail and the variability in raindrop size distribution, which entail uncertainties in the relation between reflectivity and rain rate (Germann et al., 2009; Lee et al., 2009). Several studies investigated the reliability of radar rainfall data for selected events (Gourley et al., 2010; Mandapaka et al., 2010; Rossa et al., 2010) or on common verification statistics for long samples of data (Krajewski et al., 2010; Michelson et al., 2005). However, the lack of understanding of the uncertainties in the radar estimates in the hydrological community has often led to the exclusion of radar data in hydrological modelling (Rossa et al., 2005).

Numerical weather prediction systems provide forecasts of meteorological data with a lead time up to several days. One of the most detailed NWP forecasts available in Europe is the COSMO-2 forecast from the consortium for small-scale modelling (COSMO), which provides forecasts for the next 24 hours on a mesh of 2.2 km x 2.2 km every three hours (Ament *et al.*, 2011). Through the assimilation of radar data even information on the convective scale is integrated in the model. However, assimilation, computation and dissemination of the forecast make the forecasts only available about 2.5 hours after initialisation, so that the gain from the assimilated radar data may already be lost by the time the meteorological forecast serves as input to the hydrological model. Other uncertainties involved in NWP are the uncertainties about the state of the atmosphere at the time of the model initialisation and uncertainties involved in parameterisation and downscaling (Jaun and Ahrens, 2009).

Finally there are the uncertainties directly related to the hydrological model. A model is always just a simplification of the natural system, an attempt to reproduce the natural processes, and is therefore never perfect. Apart from these uncertainties in the model structure, there are the uncertainties about the initial state of the catchment and the uncertainties in parameterisation of the model.

The overall uncertainty in a hydro-meteorological forecast system is a complex integration of all the above mentioned sources of uncertainty (Zappa *et al.*, 2011).

One way to cope with all these uncertainties is through probabilistic forecasts. Probability provides the tools to represent the precision of a forecast and this can be valuable information (Wilks, 2006).

Quantification of uncertainty

One approach to meet the problem of insufficient process description in the hydrological or meteorological model is to run several different models for the same target area. This is called a poor man's ensemble and was analysed for example by Ebert (2001), Arribas et al. (2005) or during the MAP D-PHASE project (Rotach *et al.*, 2009a; Rotach *et al.*, 2009b).

Probabilistic forecasts in the form of ensemble prediction systems (EPS) were first used among atmospheric modellers in the nineties to assess the uncertainty involved in precipitation forecasts in time and space in order to gain additional information on the characteristics of a possible event (Buizza, 2008; Molteni et al., 1996; Palmer and Buizza, 2007). In atmospheric sciences the concept of ensembles has been extended by using perturbations of the parameterisation and initial conditions of atmospheric models (Houtekamer et al., 1996). All the uncertainties in the meteorological part of the forecasting chain will sum up to the input uncertainty to the hydrological model.

The parameterisation uncertainty of the hydrological model can be met with the concept of equifinality and the Generalized Likelihood Uncertainty Estimation (GLUE) methodology (Beven and Binley, 1992). Beven and Biley (1992) suggested that since there is no perfect model structure and all observations used for model calibration are erroneous to some extent, a true parameter set describing the model cannot be expected, but equally likely parameter sets do exist. The concept of equifinality has since been applied and further developed (Beven, 2006; Beven and Freer, 2001; Choi and Beven, 2007). The advantage of ensemble forecasts is that their members are equally probable forecasts of the future state of the system. The spread of the ensemble members is therefore a quantification of the uncertainty of the forecast.

The terms nowcast and forecast

Nowcasts are generally defined as very short-term forecasts (Mandapaka et al., 2012). In the studies of this dissertation, nowcasts are forced by operationally available precipitation measurements and therefore their maximum lead time equals to the response time of the considered catchment. For predictions with lead times exceeding the internal lead time of the catchment the term forecast is used.

Objectives and Outline

NWP forecasts have a limited ability to capture the distribution of precipitation at the scale relevant for flash flood forecasting and rain-gauge networks are often scarce or absent in remote flash flood prone regions. Therefore, this dissertation focuses on the possibilities of applying weather radar data for flash flood nowcasting and forecasting and in particular in assessing the potential added value of probabilistic flash flood nowcasting and forecasting systems. The main issues that are addressed in this thesis are:

- Does the application of an ensemble technique to quantify uncertainties in radar QPE improve the performance of a flash flood early warning system?

- Is there a benefit in applying an ensemble technique to quantify uncertainties in the parameterisation of the hydrological model?

- Is there a potential benefit in using ensemble discharge nowcasts to derive ensembles of initial conditions for a subsequent forecast?

Following this introductory part of the dissertation the material and methods used in the studies described in papers I to IV are presented, a short overview of these papers is given, which is then followed by an overall synthesis and conclusion. The main part of the dissertation consist of the papers I to IV. In paper I, an extensive review on uncertainty propagation in advanced hydro-meteorological forecast systems is presented. Paper II presents a performance analysis of probabilistic and deterministic discharge nowcasts along with investigations of spread development over time of discharge nowcasts forced by a radar ensemble. In paper III, ensemble techniques to quantify uncertainties in radar QPE and model parameterisation are assessed. In paper IV, the potential of radar ensemble nowcasts to derive a set of initial conditions for a subsequent chaining of a NWP forecast is presented together with an analogue-based ensemble approach. Finally, two papers comprising additional work accomplished during the preparation of this dissertation are included in the appendix. Paper AI presents an event analysis of two severe flash floods that occurred in two small Swiss pre-alpine catchments in June 2007. Paper AII presents a case study of an operational flood forecast system that was implemented for flood protection of the city of Zürich, Switzerland.

MATERIALS AND METHODS

Hydrological Modelling

The semi-distributed rainfall-runoff model PREVAH (Precipitation Runoff Evapotranspiration HRU related model) (Gurtz et al., 2003; Viviroli et al., 2009; Zappa et al., 2003), was used in the studies carried out in papers II-IV. The model's spatial discretization relies on the delineation of hydrological response units (HRUs) that take into consideration information on topography, land use and soil depth (Gurtz et al., 1999). PREVAH was run on a spatial resolution of 500 x 500 meters and forced by hourly hydro-meteorological data. The meteorological input-variables required to run the model are air temperature, water vapour pressure, global radiation, wind speed, sunshine duration and precipitation (Viviroli et al., 2009). In case of hindcast simulations and for model calibration the variables are obtained from meteorological ground stations. For each variable a meteorological surface, with a 500 x 500 meter grid, is interpolated with an elevation detrended inverse distance weighting, while radar QPEs are downscaled to meet the spatial properties of the hydrological model (Viviroli et al., 2009).

Model calibration

The set of adjustable parameters of the PREVAH model used in this thesis originate from a default calibration for the study areas obtained from previous applications (Ranzi et al., 2007; Wöhling et al., 2006). Data of the years 1992 to 2004 was used for model calibration and verification. The year 1992 was used as initialisation period for the model, the calibration and verification period included the years 1993 to 1996 and 1997 to 2004 respectively. This calibration aims to find the parameter set that performs best in the simulation of the average flow and that has the smallest volume error between the observed and simulated time series (Viviroli et al., 2009; Zappa and Kan, 2007). In paper III additionally a Monte Carlo experiment was conducted to find equifinal model parameter sets, which then formed a parameter ensemble.

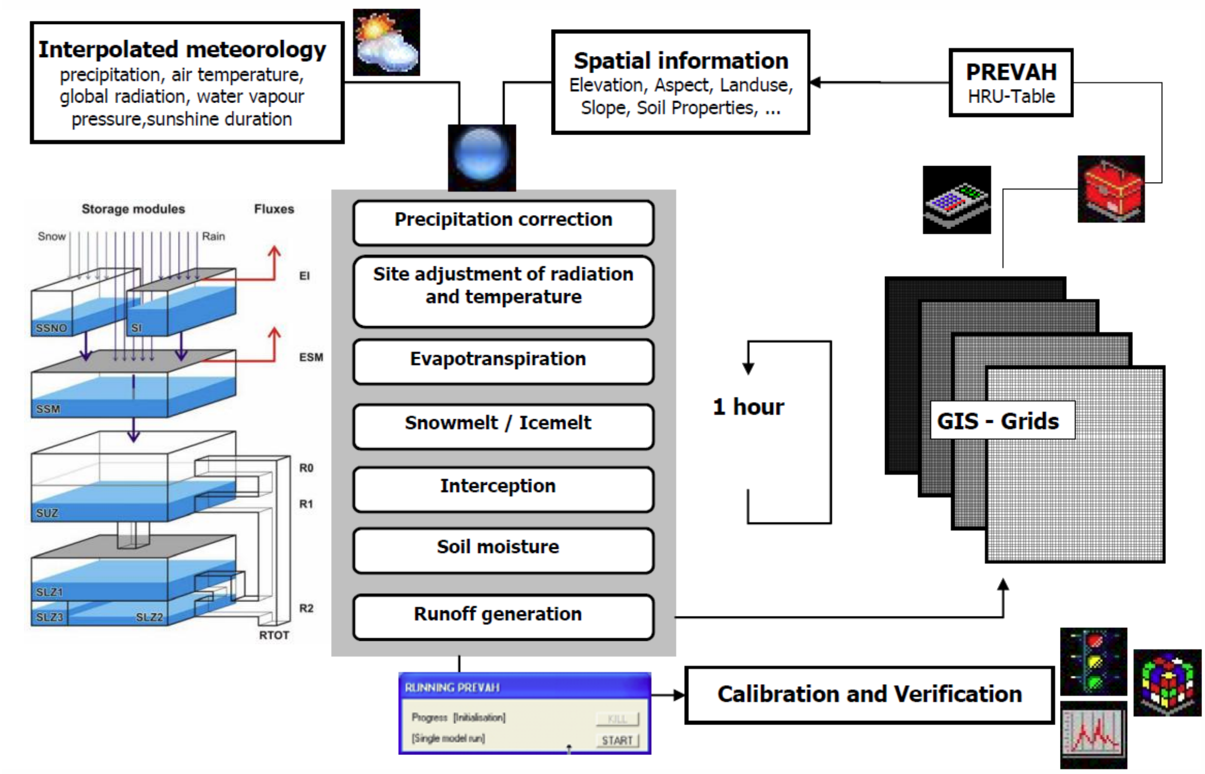

Figure 1: Structure and data flow of the PREVAH model after Viviroli et al. (2009).

Hydrological Data

All discharge data for model verification was provided by the Swiss Federal Office for the Environment (FOEN). The time series available for the five catchments investigated in Papers II-IV differ in length. Table 1 gives a summary of the available time series. The time series were also used in whole and in parts to build discharge climatologies, which served to define thresholds for the application of verification metrics.

Table 1: Catchment names, IDs, and corresponding observation records available for the studies in paper I to IV, and the design floods with a return period of two years (Q_2) according to FOEN.

Catchment	**ID**	**Observation record**	**Q_2 [m^2/s]**
Verzasca	2605	01.09.1989 to 31.12.2010	379
Pincascia	2612	01.07.1992 to 31.12.2010	99
Maggia	2368	01.01.1974 to 31.12.2010	1287
Ticino	2020	01.01.1974 to 31.12.2010	861
Calancasca	2474	01.01.1974 to 31.12.2010	141

Meteorological Data

The studies included in this dissertation evaluate the performance and applicability of hydrological forecasting chains driven with precipitation input data stemming from weather radar, ground stations and NWP. The different precipitation nowcast and forecast products and the meteorological data used for model calibration in the studies in papers II-IV are described in the following sections.

Radar quantitative precipitation information

All weather radar data used in the following experiments stem from the MeteoSwiss weather radar (Doppler, C-band) located on Monte Lema at 1650 m a.s.l., in the south of the Lago Maggiore region (Figure 3). The information on rainfall at ground level is obtained by the operational MeteoSwiss radar product for quantitative precipitation estimation (Joss and Lee, 1995). The resulting maps of quantitative precipitation estimates (QPE) have a horizontal spatial resolution of 1 x 1 km and are produced every 5 minutes. The replacement of the weather radar on Monte Lema in early summer 2011 evoke a break in the series of high quality weather radar data. Thus for the studies included in this dissertation only data until December 2010 were used.

REAL - Radar Ensemble

REAL (Radar Ensemble generator designed for usage in the Alps using LU decomposition) was developed by MeteoSwiss as a probabilistic weather radar nowcasting tool. It provides an ensemble of 25 members, each of which results from the sum of the current weather radar image and a stochastic perturbation field (Figure 2). This perturbation field is a combination of stochastic simulation techniques and detailed knowledge about the space-time variance and auto-covariance of weather radar errors (Germann *et al.*, 2009). With this methodology the residual space-time uncertainties of the atmosphere are accounted for. REAL has been produced since spring 2007 in hourly time steps and with a spatial resolution of 2 x 2 km for the Lago Maggiore region in the Southern Swiss Alps (Figure 3) (Germann *et al.*, 2009).

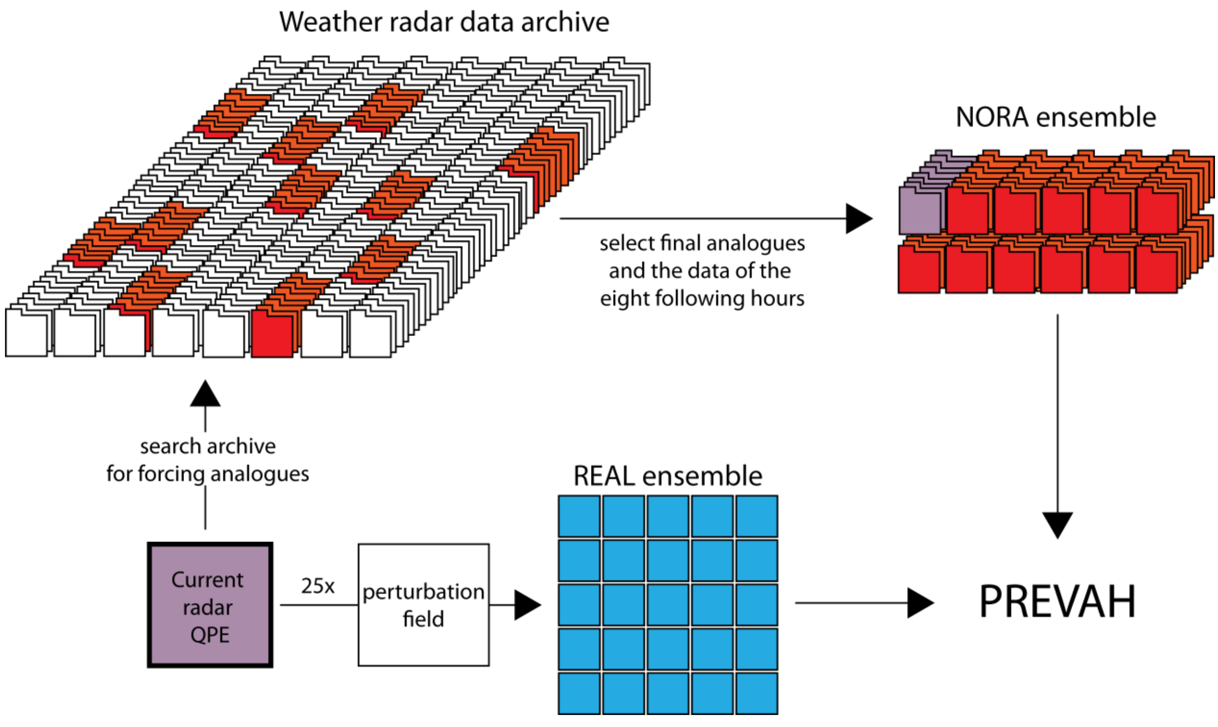

Figure 2: Procedure to build the ensembles REAL and NORA. For the REAL ensemble the current radar QPE is perturbed by a perturbation field 25 times to build an ensemble of 25 members. To build the NORA ensemble a weather radar data archive is searched for situations most similar to the current radar QPE, then those analogues and the data of the eight hours following each final analogue are extracted form the archive and build an ensemble of 12 members with 8 hours lead time each. One of these 12 members is, by construction, the Eulerian persistence of the current radar QPE (purple member of the NORA ensemble).

NORA – Nowcasting Orographic Precipitation by means of Analogues

NORA (Nowcasting of Orographic Rainfall by means of Analogues), an analogue-based heuristic nowcasting system to predict orographic rainfall for the next eight hours was developed by Panziera et al. (2011). The concept behind NORA is to find observations in the past that are most similar to the current situation, with respect to predictors describing the orographic forcing (four different mesoscale flows and air-mass stability) and two features of the radar rainfall field (wet area ratio and image mean flux). To speed up the process of finding analogues, all past weather radar data is reduced to an archive that only contains situations related to orographic forcing.

This archive was produced considering three different requirements. 1) The archive should be large enough to cover the whole range of the phenomena of interest. 2) It should be homogenous in terms of instrumental changes and data processing techniques. 3) Long-lasting and widespread events typically caused by large-scale supply of moisture towards the Alps should be selected. Isolated convection and air-mass thunderstorms were excluded from the archive. Finally all these criteria result in an archive of 71 precipitation events observed between January 2004 and December 2009, corresponding in total to 3050 hours of rainfall.

To produce the actual NORA forecast, the archive is then searched for historical situations most similar to the current one. This procedure is divided into two steps. In a first step the predictors (four mesoscale flows and air-mass stability) are used to find 120 analogues (similar situations). In a second step from these 120 analogues the 12 analogues are extracted that have the most similar rainfall pattern compared to the current one, these 12 analogues are called the final analogues. The NORA forecast is then constructed from the final analogues and the rainfall fields observed in the eight hours following each of the final analogues. This results in an ensemble of 12 members, one of which will, by construction, always be local Eulerian persistence (Figure 2). NORA is produced only if at least one of the four mesoscale winds can be estimated, otherwise no orographic forcing is expected, thus no NORA forecast is issued. A detailed description of the algorithms behind NORA is given in Panziera et al. (2011). In the experiment presented in paper IV all NORA forecasts between June 2007 and December 2010 were considered. NORA forecasts originally were issued in 5 minute time steps, but were aggregated to hourly time steps for the study to meet the setting of the hydrological model. For the past events analysed in paper IV, the whole archive was searched for analogues, this means that a hindcast of an event can also contain analogue situations that actually took place after the considered event in the past. Therefore the 24 h following the initialisation of each NORA forecast were excluded from the archive in which the analogues were sought. Panziera et al. (2011) found that results achieved in this way did not differ significantly from results when using only the hours of the archive preceding the NORA forecasts.

To date both NORA and REAL are produced by MeteoSwiss for research purposes and are only available for a test area, the Lago Maggiore region in Southern Switzerland (Figure 3).

Numerical weather prediction - COSMO-2

COSMO-2 precipitation forecast were used for the study completed in paper IV. COSMO-2 (C2) is a deterministic numerical weather prediction (NWP) model of the Consortium for Small-scale Modelling (COSMO). COSMO-2 has a lead time of 24 hours, a spatial resolution of 2.2 km and is issued every three hours (0 ,3 ,6 ,9 ,12 ,15 , 18, 21 UTC) since the beginning of the demonstration period of MAP D-PHASE (Rotach *et al.*, 2009a) in June 2007.

Ground station data

The meteorological data needed to calibrate the hydrological model PREVAH (air temperature, water vapour pressure, global radiation, sunshine duration, wind speed, and precipitation) are obtained from automatic meteorological stations on an hourly basis. The rain-gauge data used for pluviometer-based nowcasts and forecasts is composed from the automatic meteorological stations run by MeteoSwiss and from automatic rain-gauges maintained by the Canton Ticino.

All data from ground-stations are interpolated to meteorological surfaces with inverse distance weighting. For air temperature, global radiation, water vapour pressure, and wind speed, an elevation dependent trend was considered (Jaun and Ahrens, 2009; Viviroli *et al.*, 2009; Zappa and Kan, 2007) and depending on air temperature precipitation is treated as snow.

Study Catchments

The Lago Maggiore region

All the test catchments considered in the studies carried out in paper II-IV are located in the Lago Maggiore region in the Canton Ticino, Southern Switzerland. The valleys in this region are characterised by steep slopes and shallow soils with a very limited water retention capacity. The slopes are mainly covered by forests and shrubs. The discharge of the catchments of the Lago Maggiore region is governed by snow melt in spring and early summer and by heavy rainfall events in autumn. Their steep topography in combination with frequent orographic precipitation in this region make these catchments prone to flash floods and floods. All the selected test catchments are equipped with a discharge gauge at the catchment outlet (Table 1). The choice to this region is also due to the good data availability in general and also due to the availability of REAL and NORA, the new products for precipitation prediction.

The Verzasca catchment

The Verzasca catchment covers an area of 186 km^2 and elevations range from 490 to 2900 m a.s.l. The Verzasca valley is very little influenced by human activity and is not affected by water management at all down to the discharge gauge. Shortly after the gauge, the Verzasca flows into Lago di Vogorno, a retention lake for hydropower production. One rain-gauge operated by the Canton of Ticino is located inside the catchment. Within the Lago Maggiore region the Verzasca is the only gauged catchment (with respect to precipitation) not disturbed by water management. Its unspoilt state combined with the good data availability and diversity make the Verzasca very valuable for the validation of hydrological forecasting systems.

The Pincascia catchment

The Pincascia catchment, 44 km^2 in area, is a sub-catchment of the 186 km^2 Verzasca catchment. It lies in the south east of the Verzasca catchment. As the Pincascia is ungauged with respect to precipitation, results for the Pincascia can serve as an internal verification of the experiments tested in the Verzasca catchment.

The Maggia catchment

The Maggia catchment covers an area of 926 km^2 and elevations range from 200 up to 3300 m a.s.l. Since 1953 the catchment is heavily influenced by the development of a complex hydropower system in the headwaters. Due to the water management, the strong seasonal snow melt component is dampened to a rather constant base flow and occasional flood peaks. The bottom of the main valley is one of the last undisturbed floodplains in Switzerland (Perona et al., 2009).

The Ticino catchment

The Ticino catchment with its 1515 km^2 is the main catchment of the Lago Maggiore region. It is much more populated and thus more influenced by human activity than the smaller catchments of this region. The main valley of the Ticino catchment is part of one of the main traffic routes crossing the Alps. Hence the lower area of the catchment, where the valley is broad enough, is intensively used by industry and agriculture. However, the steep slopes are only little used. Altitudes range from 220 m to 3400 m asl. The influence of water management is substantial, but all water remains in the catchment and reaches the gauge in Bellinzona.

The Calancasca catchment

The Calancasca catchment covers an area of 120 km^2. The Calancasca valley is a subcatchment of the Ticino catchment, its character is very rural and alpine with steep slopes, it ranges from 740 m asl to 3200 m asl in altitude. At the top of the catchment a small glacier is situated, which accounts for 1.1% of the catchment area. The Calancasca catchment is little influenced by hydropower. The water from the headwater is partly redirected to a hydropower plant in the neighbouring catchment to the east. The water finally appears again at the gauge of the Ticino river (Figure 3). Downstream of the Calancasca gauge the stream water is stored in a small retention lake for hydropower production.

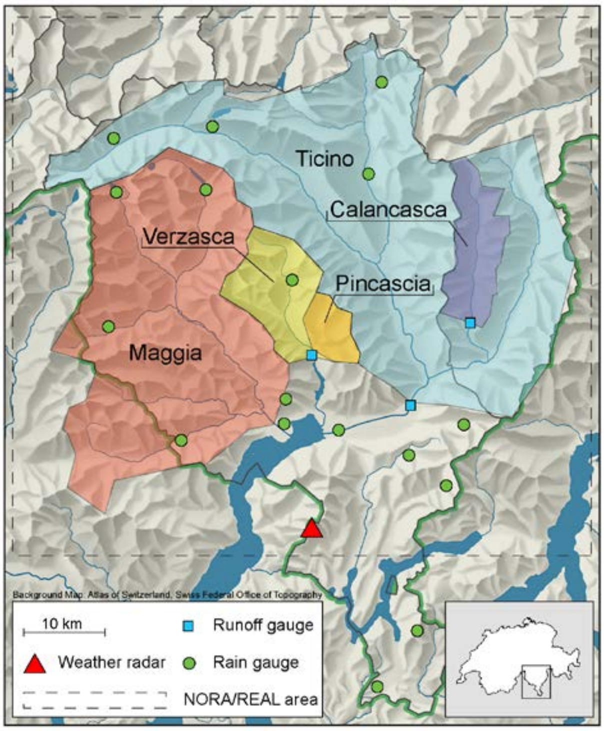

Figure 3: Lago Maggiore region with all catchments and gauging networks used in the different studies of this thesis. The dashed frame indicates the area for which REAL and NORA are available.

Verification Methods

The various nowcast and forecast experiments in the studies carried out in paper II-IV were evaluated with different verification metrics. Typically there is no ideal single skill score to express the goodness of a model simulation. Therefore the combination of several performance measures gives a better estimate of the actual performance of a model system (Bartholmes et al., 2009).

Brier Skill Score

The Brier Skill Score (BSS) is based on the Brier Score (BS). The BS is the mean squared error of the probability forecasts, given the observed outcome of an event (yes/no):

$$BS = \frac{1}{n}\sum_{d=1}^{n}\left(y_d - o_d\right)^2 , \qquad (1)$$

where o_d indicates whether the observation exceeded a predefined threshold (yes: $o_d = 1$, no: $o_d = 0$) and y_d is the forecast probability to exceed this predefined threshold ($0 \leq y_d \leq 1$).

A perfect forecast system would have a BS of zero. In order to compare the different forecast systems to each other we make use of the BSS, which sets the skill of the different forecasts in relation to a reference forecast (Wilks, 2006).

$$BSS = 1 - \frac{BS}{BS_{Cl}}, \qquad (2)$$

The reference forecast used in the studies carried out is the climatological probability to exceed the predefined discharge threshold. The definition of climatology differs between the individual studies. A perfect forecast has a skill of 1, whereas forecasts worse than the climatological forecast have a skill below 0.

False Alarm Ratio and Probability of Detection

The false alarm ratio (FAR) is the fraction of forecast threshold exceedances, that turn out to be wrong. The best FAR value is zero, which means that each positive forecast was followed by an observed threshold exceedance.

The probability of detection (POD) is the portion of correctly forecast threshold exceedances to the number of times the event really happened. The best POD value is one, which means that each observed threshold exceedance was forecast. The POD is only sensitive to missed events and not to false alarms and thus can always be improved by forecasting an event more frequently. This would, however, lead directly to an increase in false alarms and for extreme events result in an overforecasting bias (Bartholmes *et al.*, 2009; Wilks, 2006). Because of this relation of the two scores FAR and POD are always shown together in papers II-IV. Furthermore as the FAR and the POD are both skill scores derived from a 2 x 2 contingency table of prediction/observation pairs, they are tailored for non-probabilistic predictions and thus for the ensembles, scores were calculated for their median.

Rank Histograms

A Rank Histograms (RH) gives information about the spread of the an ensemble. They show where the observation is ranked in relation to all the ensemble members. A u-shaped RH indicates an underdispersed ensemble, J-shaped RHs refer to under- and overforecasting biases, a correct average spread results in a flat RH.

ROC area

The ROC area (ROCA) is the area under the ROC (relative operating characteristic) curve. A ROC curve is drawn in a ROC diagram, which incorporates information on the POD (y-axis) and the false alarm rate (x-axis) and for the whole range of forecast probabilities. The false alarm rate is the fraction of non-occurrences for which a threshold exceedance was forecast.

A perfect forecast will result in a ROC curve connecting the points (0/0), (0,1) and (1/1) of the ROC diagram. An unskilful forecast will not lie above the diagonal (0/0),(1,1). Thus the area under the ROC curve is a convenient way to express the degree of discrimination.

However, ROC is not sensitive to an overall bias, that means that ROC actually indicates potential skill, which would be achieved if the forecasts were correctly calibrated (Wilks, 2006).

A forecast system shows positive skill if the ROC area is bigger than 0.5, but according to Buizza et al. (1999) the minimum value for a forecast system to be useful for a decision maker is 0.7.

Nash Sutcliffe Efficiency

The Nash Sutcliffe efficiency (NSE) is a widely used efficiency measure in hydrology. It measures the relative improvement of the simulation compared to the mean of the observation (Schäfli and Gupta, 2007). The NSE thus penalizes errors in high flows more than errors in low flow conditions, which makes it suitable for studies that focus on high discharge.

$$NSE = 1 - \frac{\sum_{t=1}^{N} \left[q_{obs}(t) - q_{sim}(t)\right]^2}{\sum_{t=1}^{N} \left[q_{obs}(t) - \overline{q}_{obs}\right]^2} \qquad (5)$$

Where $q_{obs}(t)$ is the observed discharge at time step t, $q_{sim}(t)$ the simulated discharge and $\overline{q}_{obs}$ the mean observed discharge over the whole simulation period of length *N*. For the probabilistic nowcasts, NSE values were calculated for the ensemble median. The NSE was used in paper III to evaluate both single events and long discharge nowcast time series.

OVERVIEW OF PAPERS

Uncertainty propagation in advanced hydro-meteorological forecast systems

Paper I:

Rossa, A., Liechti, K., Zappa, M., Bruen, M., Germann, U., Haase, G., Keil, C. and Krahe, P., 2011. The COST 731 Action: A review on uncertainty propagation in advanced hydro-meteorological forecast systems. *Atmospheric Research, 100(2-3): 150-167.*

An extensive review of recent works in uncertainty propagation in hydro-meteorological forecasting systems showed that the collaboration between the meteorological and hydrological sciences communities gained strength. A bibliometric analysis by queries through the “ISI Web of Knowledge” on April 15th 2009 showed that the fields of meteorological and hydrological modelling and especially the combination of the two experienced increasing scientific attention after 2005 (Table 2). This is an encouraging finding as only through communication and knowledge transfer among the communities involved in flood forecasting can the efforts put in a full-fledged flood forecasting system pay off and its benefits be maximized.

Table 2: Bibliometric analysis by queries through the “ISI Web of Knowledge” on April 15th 2009.

(UNCERTAINTY OR PROBABILISTIC OR ENSEMBLE) AND …	First Hit	UP TO 2005		AFTER 2005		
		Papers	Cites	Recent Papers	Recent cites	Cites to recent papers (%)
HYDRO* AND MODEL*	1973	1263	10828	1324	20644	25.3
METEO* AND MODEL*	1990	438	5110	495	6354	24.7
HYDRO* AND METEO* AND MODEL*	1991	74	809	133	1451	28.5
FORECAST	1972	684	7085	757	12643	23.4

Flood nowcasting using radar ensemble

Paper II:

Liechti, K., Fundel, F., Germann, U. and Zappa, M., 2012. Flood nowcasting in the Southern Swiss Alps using radar ensemble. *In: R.J. Moore, S.J. Cole and A.J. Illingworth (Editors), Weather Radar and Hydrology. Red Book. Red Book, Exeter, UK.*

A performance analysis and comparison of radar ensemble (REAL) driven discharge nowcasts and nowcasts driven by interpolated rain-gauge data and deterministic radar QPE was assessed for daily discharge maxima. Also the development of the spread of REAL over time is evaluated. For every day in the study period from April 2007 to December 2009 a REAL nowcast was initialised and allowed to develop for ten days. Hence REAL nowcasts with identical spread development time were chained to each other resulting in ten REAL nowcast time series of 997 consecutive days each (Figure 4).

The performance analysis focused on daily discharge maxima and on thresholds corresponding to the 80% and 95% quantiles of the catchment climatologies. The ten REAL nowcast chains and the two deterministic discharge nowcasts were evaluated for the four catchments Ticino, Maggia, Verzasca and Pincascia (Figure 3) using the Brier Skill Score (BSS), the Probability of Detection (POD) and the False Alarm Ratio (FAR) (Figure 5).

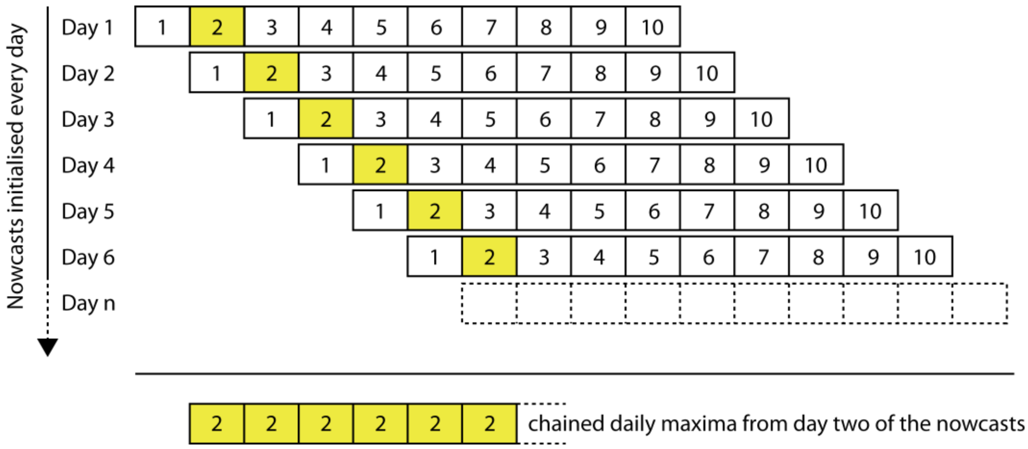

Figure 4: REAL nowcasts with identical spread development time are chained to each other. This results in 10 REAL nowcast time series for 997 consecutive days.

An optimal spread development time for the REAL nowcast could not be derived from the BSS. Decreasing performance in BSS, POD and FAR from chain 8 to 10 indicate that a spread development time longer than eight days is not useful. In all catchments REAL nowcasts showed higher skill then the deterministic radar nowcasts, while the results for the comparison of REAL nowcasts with nowcasts driven by interpolated rain-gauge data depends on the threshold chosen for the analysis.

The superiority of REAL nowcasts over the deterministic radar nowcasts is a clear argument for the probabilistic approach. Further it could be shown that REAL nowcasts provide most benefit for catchments with a sparse or no rain-gauge network. Subsequent studies focus on the use of REAL nowcasts to derive initial states for the initialisation of discharge forecasts driven by data from numerical weather prediction models (see Paper IV).

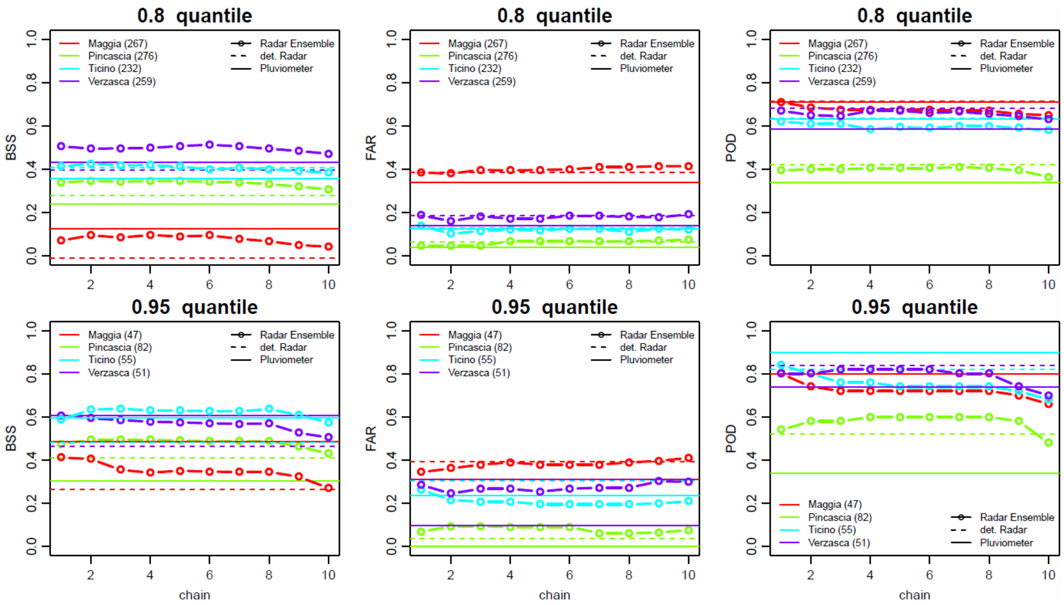

Figure 5: Brier Skill Score (BSS), False alarm ratio (FAR) and Probability of detection (POD) for the four test catchments calculated for the different precipitation input and the 0.8 and 0.95 discharge quantile. The numbers in the brackets are the number of events reaching or exceeding to the respective threshold quantile. Catchments: Maggia is heavily influenced by water management; Pincascia has no rain-gauge within the catchment.

Ensemble discharge nowcasts

Paper III:

Liechti, K., Zappa, M., Fundel, F. and Germann, U., 2012. Probabilistic evaluation of ensemble discharge nowcasts in two nested Alpine basins prone to flash floods. *Hydrological Processes: in press.*

The quality of hydrological discharge simulations depends to a great extent on the uncertainties in the meteorological input and the model parameterization. To quantify these uncertainties, ensemble techniques were tested with a four-year nowcast experiment for the Verzasca and Pincascia catchments, two nested flash-flood-prone basins in the southern Swiss Alps (Figure 3).

The spatiotemporal uncertainties in the weather radar quantitative precipitation estimates (QPE) were accounted for by applying the radar ensemble REAL (25 members). To account for uncertainties in model parameterization a Monte Carlo experiment was run to find 26 equifinal model realizations. The resulting parameter ensemble, consisting of 26 members, was run with precipitation input obtained from interpolated pluviometer data (PPE) and with the deterministic operational weather radar QPE (RPE).

To quantify the uncertainty in both the meteorological input and the model parameterization on the predictive skill of the hydrological ensemble prediction system (HEPS), the three ensemble nowcasts were compared with the corresponding deterministic versions of the experiments, i.e. nowcast driven by interpolated rain-gauge data (PLU) and nowcast driven by deterministic radar QPE (RAD).

The five discharge nowcast experiments were evaluated for a four-year time series and for single events for the Verzasca catchment and the Pincascia catchment. The performance of the different discharge nowcasts experiments were assessed with different verification metrics (BSS, POD, FAR, NSE and rank histograms).

The evaluation of the single events showed that performance of the different nowcasting experiments depends to a great extent on the characteristics of the storm. Thus no clear superiority for either pluviometer-based or radar-based nowcasts could be obtained.

However, the evaluation of the four-year nowcast experiments showed that pluviometer-based nowcasts outperform radar-based nowcasts in the gauged and calibrated Verzasca catchment and that their performance can be improved when accounting for parameterization uncertainty with a parameter ensemble (Figure 6a). For the small, ungauged (with respect to precipitation) Pincascia catchment the results achieved by the radar-based nowcasts are superior to the pluviometer-based nowcasts. Especially the radar ensemble proved to be of significant advantage for flash flood nowcasts in such catchments (Figure 6b). This result shows the potential of REAL and can serve as a motivation for potential future implementation at a larger scale or in other poorly gauged regions prone to flash floods.

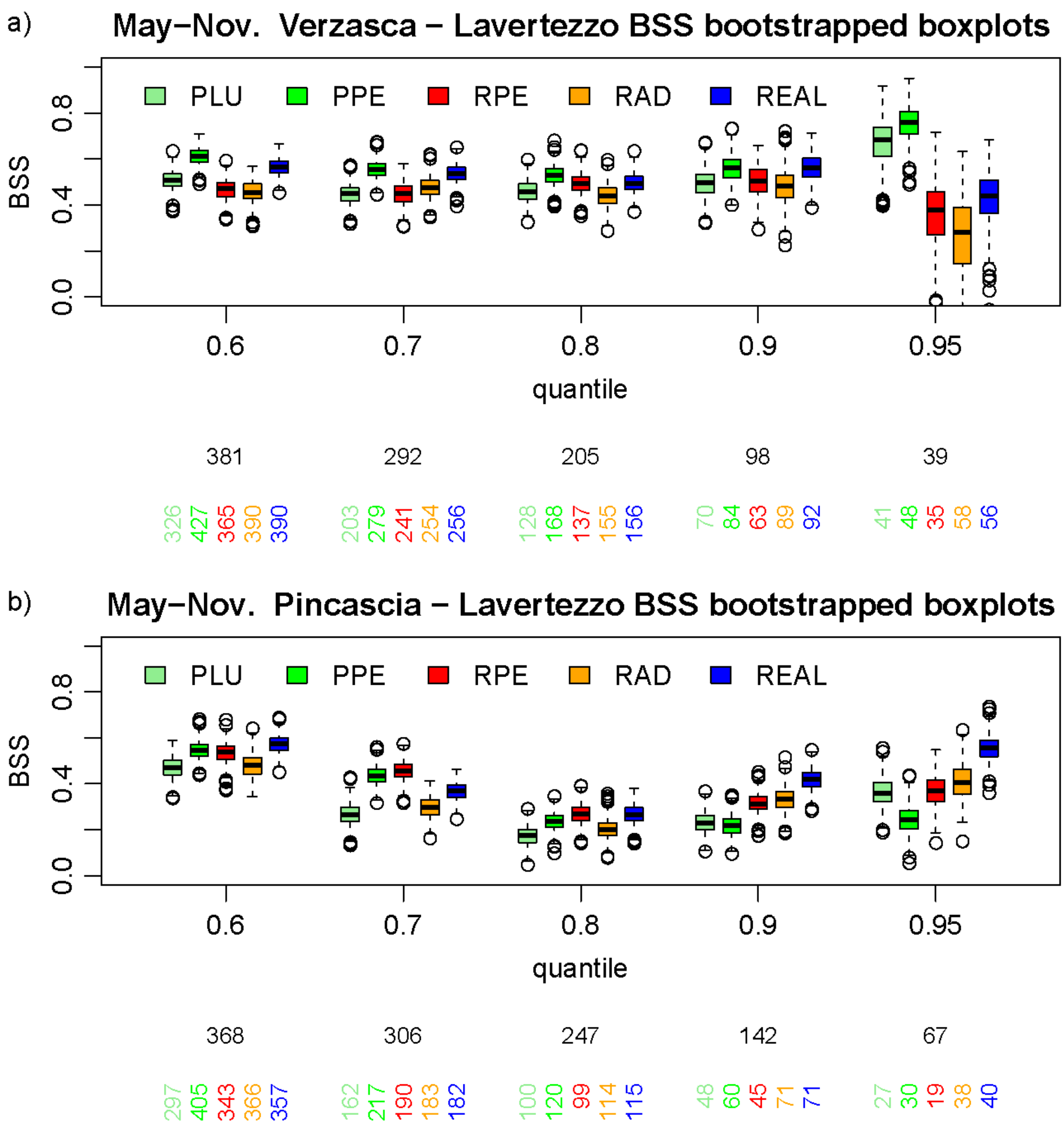

Figure 6: Brier skill scores (BSS) for different threshold quantiles (derived from the catchment climatologies) for the years 2007 to 2010 (May to November of each year). The box plots describe the distribution of BSS derived from bootstrapping the sample. Horizontal numbers below the graphs indicate the number of days where an exceedance of the threshold quantile was observed. Vertical numbers indicate for each nowcast type the number of days that were simulated to be above the threshold quantile. For ensembles, the number of ensemble members exceeding the threshold was divided by the size of the ensemble.

More potential is seen in the application of the different ensemble nowcasts to derive initial conditions for subsequent forecast experiments with longer lead times. Furthermore the predictive skill of radar-based discharge predictions could be improved if long continuous time series of weather radar data were available to calibrate hydrological models.

From radar based discharge nowcasting towards forecasting

Paper IV:

Liechti, K., Zappa, M., Panziera, L. and Germann, U. Flash flood early warning using weather radar data: from nowcasting towards forecasting. *Manuscript.*

The study presented explores the limits of radar-based forecasting for hydrological runoff prediction. Two novel probabilistic radar-based forecasting chains for flash-flood early warning are investigated in three catchments of the Lago Maggiore region (Ticino, Calancasca, Verzasca) and set in relation to deterministic discharge forecast for the same catchments.

The first probabilistic radar-based forecasting chain is driven by NORA (Nowcasting of Orographic Rainfall by means of Analogues), an analogue-based heuristic nowcasting system to predict orographic rainfall for the next eight hours. It consists of 12 members, initialized with the initial conditions derived from a four day nowcast with deterministic radar QPE. The second probabilistic forecasting system is REAL-C2, where COSMO-2 is initialized with 25 different initial conditions derived from a four day nowcast with the radar ensemble REAL. Additionally three deterministic forecasting chains were analysed. One is the persistence of the radar QPE at t0 (PERS), the other two are COSMO-2 forecasts initialised with initial conditions derived from a four day deterministic nowcast with radar QPEs (RAD-C2) and interpolated rain-gauge data (PLU-C2) respectively (Figure 7).

The performance of these five flash-flood forecasting systems were analysed for all hours between June 2007 and December 2010 for which NORA forecasts were issued, due to the presence of orographic forcing. The performance measures applied were BSS, FAR, POD, BIAS and ROC area.

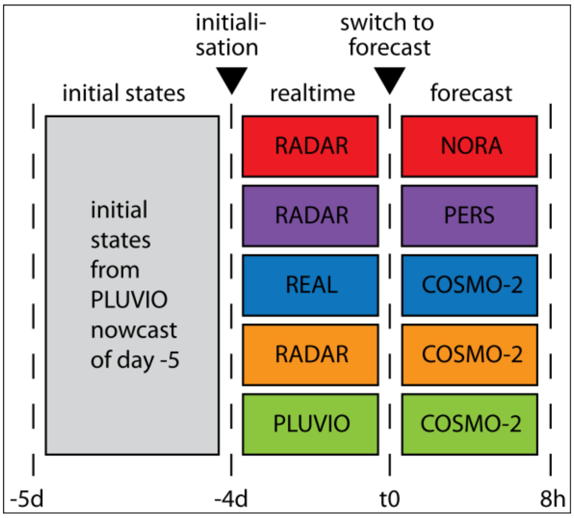

Figure 7: Scheme of the different forecasting chains.

The results show a clear preference for the probabilistic approach. NORA forecasts outperformed the PERS forecast in all catchments, over all thresholds and for all eight hours lead time. This is notable, as the events covered by NORA forecasts tend to be persistent by definition. Moreover

NORA generally outperformed RAD-C2 for thresholds up to the 80% quantile of the climatology. The best results were, however, mainly achieved by the REAL-C2 forecasting chain, which also show remarkable skill on the highest threshold. Again this shows the advantage of the probabilistic approach. Building an ensemble of 25 different initial conditions with REAL nowcasts lead to better results than NORA which starts from a single initial condition. Future investigations may use NORA forecasts to derive initial conditions for a subsequent initialisation of NWP forecasts as done in REAL-C2.

If the required data to produce REAL are available, REAL-C2 is the preferred forecasting chain, it performs better than NORA and is not restricted to events related to orographic precipitation. However, for regions where REAL cannot be produced, NORA might be an option to forecast events triggered by repetitive weather situations, like orographic forcing.

SYNTHESIS

The aim of this study was to contribute and to advance the field of flash flood forecasting. In this section the results of the different studies carried out within the framework of this dissertation are synthesised.

A review on uncertainty propagation in advanced hydro-meteorological forecast systems (paper I) suggests that a potential improvement of flash flood forecasting systems would involve efforts to improve radar QPE for small- and medium-scale river basins, and in making use of radar to improve short-range NWP. In agreement with this review, particular emphasis was given to the application of weather radar data in this dissertation. Thus, the main research questions addressed in this dissertation were:

a) Does the application of an ensemble technique to quantify uncertainties in radar QPE improve the performance of a flash flood nowcasting and forecasting system?

b) Is there a benefit in applying an ensemble technique to account for uncertainties in the parameterisation of the hydrological model?

c) Is there a potential benefit in using ensemble discharge nowcasts to derive ensembles of initial conditions for a subsequent forecast?

Concerning **a)**, it was shown that weather radar ensembles like REAL and NORA (Germann *et al.*, 2009; Panziera *et al.*, 2011) are highly valuable for flash flood early warning, especially for poorly gauged alpine catchments. In a performance analysis of discharge nowcasts for 997 days the discharge nowcasts forced by the radar ensemble REAL performed better then the discharge nowcasts forced by deterministic weather radar QPE for all four test catchments (paper II).

A more detailed analysis for two undisturbed catchments, the Verzasca and its ungauged (with respect to precipitation) subcatchment Pincascia, underline the value and potential of the radar ensemble REAL for ungauged catchments. A comparison of discharge nowcasts forced by REAL and forced by interpolated rain-gauge data showed that for the small, ungauged catchment Pincascia the discharge nowcasts forced by REAL are clearly superior to the discharge nowcasts forced by interpolated rain-gauge data. For the gauged catchment, however, this comparison has the opposite result for high discharge, and discharge nowcasts forced by interpolated rain-gauge data perform better.

Further investigations concerning **b)** showed that these deterministic nowcasts can be further improved by applying an ensemble technique to account for the parameterisation uncertainty of the hydrological model. Discharge nowcasts initialised by an ensemble of equifinal parameter sets (parameter ensemble) improved the performance of nowcasts forced by interpolated rain-gauge data for the gauged catchment Verzasca, but not for the smaller ungauged catchment Pincascia. This shows that the uncertainty in model parameterisation plays a minor role for small ungauged catchments compared to the uncertainty in the meteorological input.

These results show not only the value of radar ensembles for ungauged catchments, but also the importance and value of a dense, well-maintained rain-gauge network. For ungauged catchments, however, discharge nowcasts forced by the radar products available for these studies, seem to perform generally better or at least equally good as when forced with interpolated rain-gauge data. This is also confirmed by the event analysis discussed in paper AI. Hindcasts for two flash flood events that occurred in two small catchments (46 and 59 km^2) in the northern Swiss pre-Alps were analysed with operationally available meteorological data. Input data from radar QPE resulted in a good estimate of the flash flood in one of the catchments, whereas hindcasts with interpolated rain-gauge data failed in both catchments, also in the one with a dense network of rain-gauges. The events analysed in the different studies show that there is no universal remedy and that the performance in individual cases of flash flood events depends a lot on the characteristic of the triggering storm.

The lead time of discharge nowcasts forced by operationally available meteorological data is limited to the internal response time of the considered catchment. For small, fast responding catchments this lead time can be too short to take adequate action prior to a flash flood. Therefore several approaches to extend the lead time were analysed in paper IV. To address **c)** discharge forecasts forced by COSMO-2 forecasts were initialised with initial conditions derived from discharge nowcasts forced by interpolated rain-gauge data, radar QPE and radar ensemble REAL, respectively (paper IV). The analysis of the different experiments led to the conclusion that discharge nowcasts forced by the radar ensemble REAL are a valuable tool to derive ensembles of initial conditions for subsequent initialisation of deterministic high resolution NWP forecasts like COSMO-2, and that this approach leads to improved forecast skill compared to deterministic forecasts.

A further study investigated the feasibility to force flash flood forecasts with an analogue-based radar ensemble (NORA) tailored to forecast orographic precipitation for the next eight hours (paper IV). The forecasts based on past events performed better then the forecasts forced by the persistence of the current radar QPE. This result highlights the potential of weather radar data archives for early warning, especially for hazards related to repetitive weather situations like orographic forcing. An archive can be searched for situations similar to the present state and based on the analogue situations identified in the archive, the future development of the present situation can be assessed according to the evolution of the past analogue situations. NORA as well as REAL can be used to derive initial conditions for a subsequent coupling with a high resolution NWP forecast. The most effective way to couple forecasts forced by NORA with forecasts forced by NWP has yet to be investigated.

In all experiments conducted within the framework of this dissertation, the radar ensemble REAL performed very well, both for discharge nowcasting as well as for the generation of initial conditions for subsequent forecasts forced by NWP. The computational effort needed to produce REAL can be justified by the added value of the radar ensemble for flash flood forecasting. Therefore, it can be recommended to produce the radar ensemble REAL also for other regions, where flash flood forecasts are needed.

The prerequisite to apply the method is detailed information about the space-time variance and auto-covariance of the radar errors, which requires an adequate network of meteorological ground stations. If this prerequisite is not fulfilled, an analogue based ensemble approach like NORA can be an option, if the hazard is mainly triggered by repetitive weather situations.

Limitations and further considerations

The added value and advantages of ensemble forecasts are not necessarily evident for people not familiar with this topic. An end user may ask, where the improvement compared to the common deterministic forecast lies, when the probabilistic forecast seems to be so imprecise. It is therefore very important to communicate that probabilistic forecasts are not more imprecise than deterministic forecasts, but that the advantage is exactly that the uncertainty, which is equally inherent in deterministic forecasts, is quantified and shown, and that therefore probabilistic forecasts provide more information that can be integrated in the decision making process of the end user. To maximise the benefit of probabilistic forecasts, a close collaboration and communication between provider and user of the forecast is important.

The experiments conducted in paper II and IV also showed that the predictability of catchments influenced by water management for hydropower production is very limited with only rough assumptions made about the water retention and redirection. Efforts towards closer collaboration between hydropower companies and hydrological forecasters are therefore desirable. Both sides, operators as well as science, could profit from such a collaboration. For example, discharge forecasts for the headwaters of retention lakes can help hydropower companies to optimise their power production and avoid economical losses due to spill over, while on the other hand, detailed information about the changes in their storages and water redirections would enhance the predictability of downstream parts of the catchments. This issue is addressed in a new project, which aims to establish an operational forecasting system for the rivers of the Canton Ticino in the Lago Maggiore region. Paper AII additionally presents the set up and evaluation of a well established operational flood forecasting system for the city of Zürich that involves the close collaboration between forecasters, civil services and power plant operator.

For the experiments described in papers II to IV, the hydrological model was calibrated with precipitation input from interpolated rain-gauge data, which is also one of the reasons that the predictions forced by interpolated rain-gauge data performed remarkably well. It can be expected that the performance of predictions forced by radar QPE products would improve, if the hydrological model was calibrated with radar QPE. This, however, requires long continuous data series, which are hard to obtain due to the relatively frequent adjustments and changes of the weather radar technology.

Another issue is the time step used by the hydrological model. Operational radar QPE are available every five minutes, and therefore it could be argued that, at least for discharge nowcasts, use should be made of this information by running the hydrological model at a higher temporal resolution. If this additional computational effort is worthwhile, would still have to be tested.

Another approach to forecast flash floods in ungauged catchments could then also be to test simpler models that need less information or assumptions about the characteristics and state of the catchment than a semi-distributed model like PREVAH. One example for such a model is the single-equation rainfall-runoff model presented by Kirchner (2009). It would have to be investigated, if a simpler model tailored for flash flood forecasting would fit the purpose for poorly gauged or ungauged catchments.

One of the main limitations that is met in many studies analysing the performance of flood or flash flood forecast systems , including also the studies in paper II to IV of this dissertation, is that only few big events are included in the time series available for analysis. Therefore, the forecasting systems are tested on relatively low levels, corresponding to low return periods, which leaves open the question as to what extent the conclusions drawn from the analysis are valid for extreme events.

CONCLUSIONS

- Radar ensembles enhance the performance of flash flood early warning systems, especially for poorly gauged catchments.

- Forecasts can be improved by applying a parameter ensemble. For ungauged catchments, however, parameter uncertainty plays a minor role compared to the meteorological input uncertainty.

- Flash flood forecasts forced by high resolution NWP forecasts benefit from being chained to probabilistic nowcasts.

The studies conducted within the framework of this dissertation demonstrated the value of weather radar products for ungauged catchments and the advantage of probabilistic over deterministic flash flood early warning systems.

REFERENCES

Ament, F., Weusthoff, T. and Arpagaus, M., 2011. Evaluation of MAP D-PHASE heavy precipitation alerts in Switzerland during summer 2007. Atmospheric Research, 100(2-3): 178-189.

Arribas, A., Robertson, K.B. and Mylne, K.R., 2005. Test of a poor man's ensemble prediction system for short-range probability forecasting. Monthly Weather Review, 133(7): 1825-1839.

Barredo, J.I., 2007. Major flood disasters in Europe: 1950-2005. Natural Hazards, 42(1): 125-148.

Bartholmes, J.C., Thielen, J., Ramos, M.H. and Gentilini, S., 2009. The european flood alert system EFAS – Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts. Hydrology and Earth System Sciences, 13(2): 141-153.

Beven, K., 2006. A manifesto for the equifinality thesis. Journal of Hydrology, 320(1-2): 18-36.

Beven, K. and Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes, 6(3): 279-298.

Beven, K. and Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology, 249(1-4): 11-29.

Borga, M., Anagnostou, E.N., Blöschl, G. and Creutin, J.D., 2011. Flash flood forecasting, warning and risk management: the HYDRATE project. Environmental Science & Policy, 14(7): 834-844.

Borga, M., Boscolo, P., Zanon, F. and Sangati, M., 2007. Hydrometeorological analysis of the 29 August 2003 flash flood in the Eastern Italian Alps. Journal of Hydrometeorology, 8(5): 1049-1067.

Buizza, R., 2008. The value of probabilistic prediction. Atmospheric Science Letters, 9(2): 36-42.

Buizza, R., Hollingsworth, A., Lalaurette, E. and Ghelli, A., 1999. Probabilistic predictions of precipitation using the ECMWF ensemble prediction system. Weather and Forecasting, 14(2): 168-189.

Chiang, Y.-M., Hsu, K.-L., Chang, F.-J., Hong, Y. and Sorooshian, S., 2007. Merging multiple precipitation sources for flash flood forecasting. Journal of Hydrology, 340(3-4): 183-196.

Choi, H.T. and Beven, K., 2007. Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework. Journal of Hydrology, 332(3-4): 316-336.

Collier, C.G., 2007. Flash flood forecasting: What are the limits of predictability? Quarterly Journal of the Royal Meteorological Society, 133(622): 3-23.

Delrieu, G., Braud, I., Berne, A., Borga, M., Boudevillain, B., Fabry, F., Freer, J., Gaume, E., Nakakita, E., Seed, A., Tabary, P. and Uijlenhoet, R., 2009. Weather radar and hydrology Preface. Advances in Water Resources, 32(7): 969-974.

Ebert, E.E., 2001. Ability of a poor man's ensemble to predict the probability and distribution of precipitation. Monthly Weather Review, 129(10): 2461-2480.

Gaume, E., Bain, V., Bernardara, P., Newinger, O., Barbuc, M., Bateman, A., Blaskovicova, L., Bloschl, G., Borga, M., Dumitrescu, A., Daliakopoulos, I., Garcia, J., Irimescu, A., Kohnova, S., Koutroulis, A., Marchi, L., Matreata, S., Medina, V., Preciso, E., Sempere-Torres, D., Stancalie, G., Szolgay, J., Tsanis, I., Velasco, D. and Viglione, A., 2009. A compilation of data on European flash floods. Journal of Hydrology, 367(1-2): 70-78.

Germann, U., Berenguer, M., Sempere-Torres, D. and Zappa, M., 2009. REAL - Ensemble radar precipitation estimation for hydrology in a mountainous region. Quarterly Journal of the Royal Meteorological Society, 135(639): 445-456.

Germann, U., Galli, G., Boscacci, M. and Bolliger, M., 2006. Radar precipitation measurement in a mountainous region. Quarterly Journal of the Royal Meteorological Society, 132(618): 1669-1692.

Gourley, J.J., Giangrande, S.E., Hong, Y., Flamig, Z.L., Schuur, T. and Vrugt, J.A., 2010. Impacts of Polarimetric Radar Observations on Hydrologic Simulation. Journal of Hydrometeorology, 11(3): 781-796.

Griffiths, P.G., Magirl, C.S., Webb, R.H., Pytlak, E., Troch, P.A. and Lyon, S.W., 2009. Spatial distribution and frequency of precipitation during an extreme event: July 2006 mesoscale convective complexes and floods in southeastern Arizona. Water Resource Research, 45(W07419): doi:10.1029/2008WR007380.

Gurtz, J., Baltensweiler, A. and Lang, H., 1999. Spatially distributed hydrotope-based modelling of evapotranspiration and runoff in mountainous basins. Hydrological Processes, 13(17): 2751-2768.

Gurtz, J., Zappa, M., Jasper, K., Lang, H., Verbunt, M., Badoux, A. and Vitvar, T., 2003. A comparative study in modelling runoff and its components in two mountainous catchments. Hydrological Processes, 17(2): 297-311.

Hapuarachchi, H.A.P., Wang, Q.J. and Pagano, T.C., 2011. A review of advances in flash flood forecasting. Hydrological Processes, 25(18): 2771-2784.

Hilker, N., Badoux, A. and Hegg, C., 2009. The Swiss flood and landslide damage database 1972-2007. Natural Hazards and Earth System Sciences, 9(3): 913-925.

Houtekamer, P.L., Lefaivre, L., Derome, J., Ritchie, H. and Mitchell, H.L., 1996. A system simulation approach to ensemble prediction. Monthly Weather Review, 124(6): 1225-1242.

Jaun, S. and Ahrens, B., 2009. Evaluation of a probabilistic hydrometeorological forecast system. Hydrology and Earth System Sciences, 13(7): 1031-1043.

Joss, J. and Lee, R., 1995. THE APPLICATION OF RADAR-GAUGE COMPARISONS TO OPERATIONAL PRECIPITATION PROFILE CORRECTIONS. Journal of Applied Meteorology, 34(12): 2612-2630.

Kirchner, J.W., 2009. Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. Water Resources Research, 45.

Krajewski, W.F. and Smith, J.A., 2002. Radar hydrology: rainfall estimation. Advances in Water Resources, 25(8-12): 1387-1394.

Krajewski, W.F., Villarini, G. and Smith, J.A., 2010. RADAR-RAINFALL UNCERTAINTIES Where are We after Thirty Years of Effort? Bulletin of the American Meteorological Society, 91(1): 87-94.

Lee, C.K., Lee, G., Zawadzki, I. and Kim, K.E., 2009. A Preliminary Analysis of Spatial Variability of Raindrop Size Distributions during Stratiform Rain Events. Journal of Applied Meteorology and Climatology, 48(2): 270-283.

Mandapaka, P.V., Germann, U., Panziera, L. and Hering, A., 2012. Can Lagrangian Extrapolation of Radar Fields be used for Precipitation Nowcasting over Complex Alpine Orography? Weather and Forecasting, 27(1): 28-49.

Mandapaka, P.V., Villarini, G., Seo, B.-C. and Krajewski, W.F., 2010. Effect of radar-rainfall uncertainties on the spatial characterization of rainfall events. Journal of Geophysical Research-Atmospheres, 115: 16.

Michelson, D., Einfalt, T., Holleman, I., Gjertsen, U., Friedrich, K., Haase, G., Lindskog, M. and Jurczyk, A., 2005. Weather Radar Data Quality in Europe: Quality Control and Characterization, COST 717 document, Brussels, pp. 87.

Michelson, D.B., 2004. Systematic correction of precipitation gauge observations using analyzed meteorological variables. Journal of Hydrology, 290(3â€“4): 161-177.

Molteni, F., Buizza, R., Palmer, T.N. and Petroliagis, T., 1996. The ECMWF ensemble prediction system: Methodology and validation. Quarterly Journal of the Royal Meteorological Society, 122(529): 73-119.

Montani, A., Cesari, D., Marsigli, C. and Paccagnella, T., 2011. Seven years of activity in the field of mesoscale ensemble forecasting by the COSMO-LEPS system: main achievements and open challenges. Tellus Series a-Dynamic Meteorology and Oceanography, 63(3): 605-624.

Morin, E., Jacoby, Y., Navon, S. and Bet-Halachmi, E., 2009. Towards flash-flood prediction in the dry Dead Sea region utilizing radar rainfall information. Advances in Water Resources, 32(7): 1066-1076.

Palmer, T.N. and Buizza, R., 2007. Fifteenth anniversary of EPS. ECMWF Newsletter, 114: 14.

Panziera, L. and Germann, U., 2010. The relation between airflow and orographic precipitation on the southern side of the Alps as revealed by weather radar. Quarterly Journal of the Royal Meteorological Society, 136(646): 222-238.

Panziera, L., Germann, U., Gabella, M. and Mandapaka, P.V., 2011. NORA–Nowcasting of Orographic Rainfall by means of Analogues. Quarterly Journal of the Royal Meteorological Society, 137(661): 2106-2123.

Perona, P., Molnar, P., Savina, M. and Burlando, P., 2009. An observation-based stochastic model for sediment and vegetation dynamics in the floodplain of an Alpine braided river. Water Resources Research, 45: 13.
Perry, C.A., 2000. Significant floods in the United States during the 20th Century – USGS measures a century of floods. Fact Sheet 024-00.
Ranzi, R., Zappa, M. and Bacchi, B., 2007. Hydrological aspects of the Mesoscale Alpine Programme: Findings from field experiments and simulations. Quarterly Journal of the Royal Meteorological Society, 133(625): 867-880.
Rossa, A., Bruen, M., Fruehwald, D., Macpherson, B., Holleman, I., Michelson, D. and Michaelides, S., 2005. COST 717 Action - Use of Radar Observation in Hydrology and NWP Models. COST Meteorology, EUR 21954. COST, 286 pp.
Rossa, A., Laudanna Del Guerra, F., Borga, M., Zanon, F., Settin, T. and Leuenberger, D., 2010. Radar-driven High-resolution Hydro-meteorological Forecasts of the 26 September 2007 Venice Flash Flood. Journal of Hydrology, 394(1-2): 230-240.
Rotach, M.W., Ambrosetti, P., Ament, F., Appenzeller, C., Arpagaus, M., Bauer, H.-S., Behrendt, A., Bouttier, F., ois, Buzzi, A., Corazza, M., Davolio, S., Denhard, M., Dorninger, M., Fontannaz, L., Frick, J., Fundel, F., Germann, U., Gorgas, T., Hegg, C., Hering, A., Keil, C., Liniger, M.A., Marsigli, C., McTaggart-Cowan, R., Montaini, A., Mylne, K., Ranzi, R., Richard, E., Rossa, A., Santos, M., oz, D., Sch, r, C., Seity, Y., Staudinger, M., Stoll, M., Volkert, H., Walser, A., Wang, Y., Werhahn, J., Wulfmeyer, V. and Zappa, M., 2009a. MAP D-PHASE: Real-Time Demonstration of Weather Forecast Quality in the Alpine Region. Bulletin of the American Meteorological Society, 90(9): 1321-1336.
Rotach, M.W., Ambrosetti, P., Appenzeller, C., Arpagaus, M., Fontannaz, L., Fundel, F., Germann, U., Hering, A., Liniger, M.A., Stoll, M., Walser, A., Ament, F., Bauer, H.-S., Behrendt, A., Wulfmeyer, V., Bouttier, F.o., Seity, Y., Buzzi, A., Davolio, S., Corazza, M., Denhard, M., Dorninger, M., Gorgas, T., Frick, J., Hegg, C., Zappa, M., Keil, C., Volkert, H., Marsigli, C., Montaini, A., McTaggart-Cowan, R., Mylne, K., Ranzi, R., Richard, E., Rossa, A., Santos-MuÃ±oz, D., SchÃ¤r, C., Staudinger, M., Wang, Y. and Werhahn, J., 2009b. Supplement to MAP D-PHASE: Real-Time Demonstration of Weather Forecast Quality in the Alpine Region: Additional Applications of the D-Phase Datasets. Bulletin of the American Meteorological Society, 90(9): S28-S32.
Rusjan, S., Kobold, M. and Mikoš, M., 2009. Characteristics of the extreme rainfall event and consequent flash floods in W Slovenia in September 2007. Natural Hazards and Earth System Sciences, 9(3): 947-956.
Schiemann, R., Liniger, M.A. and Frei, C., 2010. Reduced space optimal interpolation of daily rain gauge precipitation in Switzerland. J. Geophys. Res., 115(D14109): 18.
Sevruk, B., 1996. Adjustment of tipping-bucket precipitation gauge measurements. Atmospheric Research, 42(1â€“4): 237-246.
Smith, R.B., 1979. The Influence of Mountains on the Atmosphere. In: B. Saltzman (Editor), Advances in Geophysics. Elsevier, New York, pp. 87-230.
Thornton, P.E., Running, S.W. and White, M.A., 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. Journal of Hydrology, 190(3-4): 214-251.
Tobin, C., Nicotina, L., Parlange, M.B., Berne, A. and Rinaldo, A., 2011. Improved interpolation of meteorological forcings for hydrologic applications in a Swiss Alpine region. Journal of Hydrology, 401(1-2): 77-89.
USGS, 2008. Flood Definitions. U.S. Geological Survey, Kansas Water Science Center, available at: http://ks.water.usgs.gov/waterwatch/flood/definition.html.
Velasco-Forero, C.A., Sempere-Torres, D., Cassiraga, E.F. and Jaime GÃ³mez-HernÃ¡ndez, J., 2009. A non-parametric automatic blending methodology to estimate rainfall fields from rain gauge and radar data. Advances in Water Resources, 32(7): 986-1002.
Viviroli, D., Zappa, M., Gurtz, J. and Weingartner, R., 2009. An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools. Environmental Modelling & Software, 24(10): 1209-1222.
Werner, M. and Cranston, M., 2009. Understanding the Value of Radar Rainfall Nowcasts in Flood Forecasting and Warning in Flashy Catchments. Meteorological Applications, 16(1): 41-55.

Wilks, D.S., 2006. Statistical methods in the atmospheric sciences. Elsevier, Amsterdam, 627 pp.
Wöhling, T., Lennartz, F. and Zappa, M., 2006. Technical Note: Updating procedure for flood forecasting with conceptual HBV-type models. Hydrology and Earth System Sciences, 10(6): 783-788.
Zappa, M., Jaun, S., Germann, U., Walser, A. and Fundel, F., 2011. Superposition of three sources of uncertainties in operational flood forecasting chains. Atmospheric Research, 100(2-3): 246-262.
Zappa, M. and Kan, C., 2007. Extreme heat and runoff extremes in the Swiss Alps. Natural Hazards and Earth System Sciences, 7(3): 375-389.
Zappa, M., Pos, F., Strasser, U., Warmerdam, P. and Gurtz, J., 2003. Seasonal water balance of an Alpine catchment as evaluated by different methods for spatially distributed snowmelt modelling. Nordic Hydrology, 34(3): 179-202.

ACKNOWLEDGEMENTS

Frankly, doing a PhD was not on my list four years ago. But then things took their course… This very interesting European project IMPRINTS on IMproving Preparedness and RIsk maNagemenT for flash flood and debriS flow events came up. Massimiliano gave me the opportunity to do a PhD within this project, Manfred convinced me that doing a PhD is not too bad (with the backdoor open, that I could still quit after one year in case I don't like it) and Massimiliano had the patience to wait for my decision. As you may guess, my decision was positive, I did not quit after one year and I did not regret my decision. So a first big thank you goes to Massimiliano Zappa, my supervisor, for his patience, professional scientific support and inspirational ideas and inputs. Also, one can learn a lot from Massimiliano when it comes to working efficiently. I know few people as good at optimising the "cost-benefit" equation of work.

During my years at the WSL I met many nice people many of whom have become good friends. I would like to especially thank members of the group "Mountain Hydrology and Mass Movements" for their support and all the nice discussions over coffee or lunchtime. It was the good atmosphere at work that always gave me the extra motivation.

Thanks also to the IMPRINTS project partners for fruitful scientific discussions. IMPRINTS gave me the opportunity to gain insight into the structure and dynamics of a big international research project. Thanks to the MeteoSwiss team, our closest project partner, whose infectious enthusiasm was always a source of motivation.

Special thanks to Jan Seibert for accepting and supporting me as an external PhD student. Even if I was a seldom visitor at the Institute of Geography, I always felt welcome. Thanks also to Micha Werner, who agreed to be the external reviewer and examiner of my thesis. Thanks to the whole H2K group for integrating me in all kinds of activities. I hope we will have some more "running water" events together.

The biggest thanks go to my family and friends. My parents always supported me and gave me the calm needed and a good feeling for the essentials in life, along with my grandmother whose sense of humour and pragmatism will always keep me down to earth.

Dear friends, thank you all for being there, and for all your supportive distraction throughout the years, I am looking forward to more!

CURRICULUM VITAE

Personal Information

Name	Katharina Liechti
Date and place of birth	April 8 1982, Affoltern am Albis, Switzerland
Citizenship	Landiswil (BE), Switzerland

Education

04/2009 – 12/2012	Ph.D. student at the Swiss Federal Institute for Forest, Snow and Landscape Research (WSL) and the Department of Geography of the University of Zurich (Switzerland). Thesis title: *Uncertainty Propagation in Operational Flash Flood Forecasting Systems.*
09/2006 - 11/2008	Master studies in Geography at the University of Zurich. Specialisation: Physical Geography, Remote Sensing and Geographic Information Sciences. Masterthesis: *Starke Gewitter im Juni 2007, Einordnung und Hydrologische Modellierung für die Regionen Huttwil (BE) und Einsiedeln (SZ).*
08/2006 – 03/2007	ERASMUS at Uppsala University, Sweden.
10/2003 – 06/2006	Bachelor studies in Geography at the University of Zurich. Specialisation: Physical Geography, Remote Sensing and Geographic Information Sciences. Minor subjects: Geology, Mathematics, Cartography.
1998 – 2002	Matura at Kantonsschule Limmattal, Urdorf, Switzerland. New Languages Profile (English, French, Italian) with additional focus on Mathematics.

I

Atmospheric Research 100 (2011) 150–167

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmos

The COST 731 Action: A review on uncertainty propagation in advanced hydro-meteorological forecast systems

Andrea Rossa [a,*], Katharina Liechti [b], Massimiliano Zappa [b], Michael Bruen [c], Urs Germann [d], Günther Haase [e], Christian Keil [f,g], Peter Krahe [h]

[a] *ARPAV, Centro Meteorologico Teolo, Italy*
[b] *Swiss Federal Research Institute WSL, Birmensdorf, Switzerland*
[c] *Centre for Water Resources Research, University College Dublin, Ireland*
[d] *Swiss Federal Office of Meteorology and Climatology MeteoSwiss, Switzerland*
[e] *Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden*
[f] *Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany*
[g] *Meteorologisches Institut, Ludwig-Maximilians-Universität München, München, Germany*
[h] *Bundesanstalt für Gewässerkunde, Koblenz, Germany*

A R T I C L E I N F O

Article history:
Received 18 June 2009
Received in revised form 13 November 2010
Accepted 16 November 2010

Keywords:
Uncertainty
Flood forecasting
Radar
NWP
EPS
COST
MAP D-PHASE

A B S T R A C T

Quantifying uncertainty in flood forecasting is a difficult task, given the multiple and strongly non-linear model components involved in such a system. Much effort has been and is being invested in the quest of dealing with uncertain precipitation observations and forecasts and the propagation of such uncertainties through hydrological and hydraulic models predicting river discharges and risk for inundation. The COST 731 Action is one of these and constitutes a European initiative which deals with the quantification of forecast uncertainty in hydro-meteorological forecast systems. COST 731 addresses three major lines of development: (1) combining meteorological and hydrological models to form a forecast chain, (2) propagating uncertainty information through this chain and make it available to end users in a suitable form, (3) advancing high-resolution numerical weather prediction precipitation forecasts by using non-conventional observations from, for instance, radar to determine details in the initial conditions on scales smaller than what can be resolved by conventional observing systems. Recognizing the interdisciplinarity of the challenge COST 731 has organized its work forming Working Groups at the interfaces between the different scientific disciplines involved, i.e. between observation and atmospheric (and hydrological) modelling (WG-1), between atmospheric and hydrologic modelling (WG-2) and between hydrologic modelling and end-users (WG-3).

This paper summarizes the COST 731 activities and its context, provides a review of the recent progress made in dealing with uncertainties in flood forecasting, and sets the scene for the papers of this Thematic Issue. In particular, a bibliometric analysis highlights the strong recent increase in addressing the uncertainty analysis in flood forecasting from an integrated perspective. Such a perspective necessarily involves the area of meteorology, hydrology, and decision making in order to take operational advantage of the scientific progress, an aspect in which COST 731 is successfully contributing to furthering the flood damage mitigation capabilities in Europe.

1. Introduction

Floods are among the most commonly occurring types of natural disasters in Europe and their frequency and people's

* Corresponding author. ARPAV, Centro Meteorologico Teolo, Via Marconi 55, I-35037 TEOLO (PD), Italy.
E-mail address: arossa@arpa.veneto.it (A. Rossa).

0169-8095/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.atmosres.2010.11.016

vulnerability is increasing across Europe (Barredo, 2007). This is because of increased development pressures on floodplains with more and more people and valuable infrastructure moving into flood-prone areas. Thus an increase in flood-damages in Europe is foreseeable, even without taking climate change into account, (Mitchell, 2003). The Munich Re interactive system NATHAN (natural hazard assessment network, see www.munichre.com) contains an interactive map and thematic information on more than 140 major flood events in Europe from 1900 to the present, including data on casualties and economic losses. The European Commission recognized the paramount importance of the natural hazards issue for the protection of the environment and the citizens and made significant investments in associated research and development ever since the 4th Framework Programme launched in 1994. One particular focus is the real-time forecasting of extreme events with significant flooding potential, which is the basis for triggering a range of mitigating actions. This is a difficult task for many reasons, a main one being related to successfully integrating the contribution of the different 'spheres' (atmosphere, hydrosphere) each of which with its different modelling approaches, inherent uncertainties, limitations and, not least, paradigms of application.

Uncertainty recognizes a certain amount of fuzziness in the forecasts which contrasts with the definitive, categorical, decisions that flood relief managers have to take in order to launch mitigating actions. Significant effort is needed towards reconciling this apparent clash of paradigms by working on the conceptual and communication issues between scientists and decision makers related to uncertainty information (e.g. Demeritt et al., 2007). Such clarification is necessary even on a purely scientific level, in that uncertainty information seems to be perceived fundamentally differently by weather forecasters and flood forecasters.

The COST 731 Action can be seen as the expression of the will of a large number of European meteorological and hydrological services to further both understanding and, even more so, application of systematic uncertainty information. It hence focuses on hydro-meteorological forecasting and how to deal with the uncertainties inherent in the entire forecast chain. The COST 731 Action was proposed and launched in mid 2005 for a five-year period as an offspring of a series of COST Actions related to radar meteorology (Rossa et al., 2005a). While the COST Actions 72, 73, and 75 dealt with pure scientific issues related to single weather radars, radar networking, and advanced radar capabilities (Meischner et al., 1997), Action 717 focused on the application of radars in hydrological and NWP models (Rossa et al., 2005b). Here, the unparalleled ability of the radar to observe precipitation in $3+1$ dimensions was explored for validating NWP precipitation forecasts and to improve the model's initial conditions. In addition, COST 717 sought to promote the use of radar quantitative precipitation estimates (QPE) for hydrological modelling. On a non-technical level, two issues stood out: the need for a clearer communication between the participating scientific groups, and the necessity to quantify, or at least describe, the variable quality of the radar-derived QPE. The lack of understanding of this uncertainty often led to the exclusion of these data. COST 731 was, therefore, designed to address the quantification and communication of the uncertainty in meteorological observation and forecasting along with their effect on hydrological forecasting, and the subsequent impact on the decision making process.

Full-fledged flood forecasting systems which make use of meteorological forecasts to extend warning lead times are relatively recent but many operational centers around the world are now increasingly moving towards such systems (e.g. Cloke and Pappenberger 2009). In these, test beds are often developed and play an important role in exploring the potential of integrated probabilistic flood forecasting systems (Schaake et al., 2007b; Rotach et al., 2009). Also, they are indispensable opportunities for gathering hands-on experience and provide training for operational staff, without which it will be very hard to resolve the communication and paradigm difficulties. As an illustration for this trend Fig. 1 shows a COST 731 test bed example which consists in a real-time implementation of the semi-distributed rainfall–runoff model PREVAH (Viviroli et al., 2009) for flash flood modelling in a steep Alpine catchment with a probabilistic radar rainfall input. It is one of the first experiments of its kind worldwide producing operational ensemble runoff nowcasting.

In this paper some recent advances in dealing with uncertainties in flood forecasting systems are reviewed to provide the scientific context for COST Action 731 'Propagation of Uncertainty in Advanced Meteo-Hydrological Forecast Systems'. In Section 2 the organization and strategy of COST 731 is described and a bibliometric analysis of the scientific and operational relevance of the topic is presented, while Section 3 contains the main scientific review. Then the major emerging results of the Action are given in Section 4, before closing with a short summary and outlook in the final section.

2. Scientific context and operational relevance of COST Action 731

The main goal of the COST 731 Action can be summarized as quantifying, reducing, and propagating uncertainty in advanced hydro-meteorological forecast systems. In this section the context for COST 731 is given in terms of how this subject has recently gained momentum in the international scientific community, important research and demonstration initiatives as MAP D-PHASE, HEPEX and a number of European Framework Programme projects, and basic considerations on conceptual and operational implications of having to deal with uncertainty. Also goals and structure of the COST 731 Action are described.

2.1. Bibliometric analysis

It is important for a COST Action to deal with topics which are currently highly relevant within the scientific and/or the operational community. An increasing number of analyses based on bibliometric information, obtained from publication databases such as the "ISI Web of Knowledge" (http://www.isiknowledge.com/) or "Google-Scholar" (http://scholar.google.com/), are used to determine if certain research topics are "hot", i.e. are active research areas. Such analyses depend, to a certain extent, on the subjective choice of the terms for querying the databases and the criteria used for refining the "hits" in order to eliminate irrelevant publications. One approach could be to compare the hits given by a series of slight variations on a

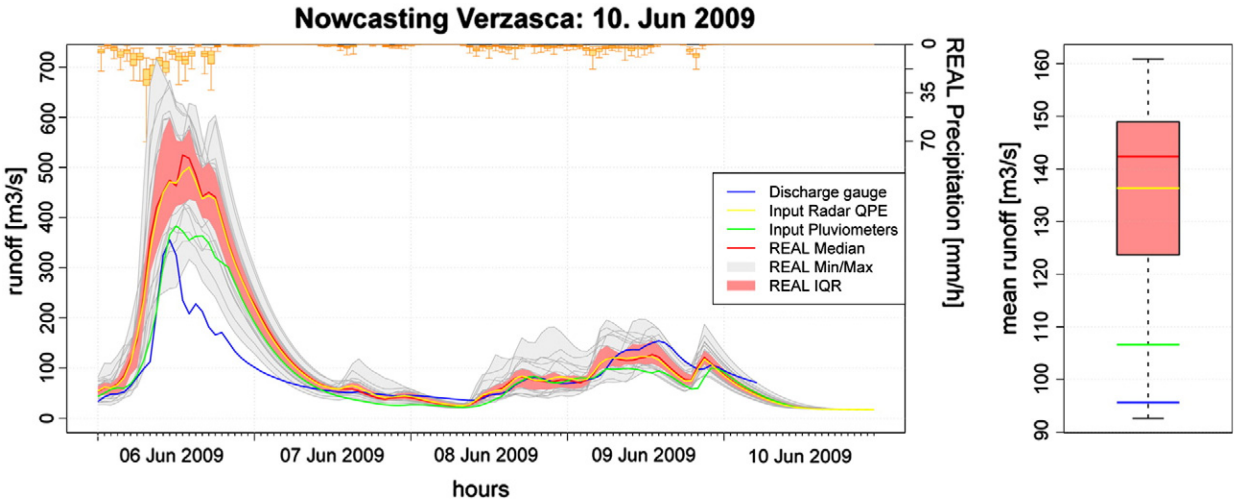

Fig. 1. Hydrological ensemble nowcasting with REAL and PREVAH, starting on 6 June 2009 for the Verzasca basin in southern Switzerland. The 25 members from REAL (light grey) are shown with corresponding interquartile range (REAL IQR, red area) and the median (red line). Additionally, two deterministic runs are shown: deterministic radar QPE (yellow line) and forcing with interpolated pluviometer data (green line). The observed runoff is shown in blue. The left panel shows the hydrograph from the initialization point up to June 10, 2009. Spatially interpolated observed precipitation as ensemble precipitation from the REAL members (orange whisker-plots). The right panel shows the average runoff between June 6th and June 9th 2009.

specific query and this might provide a more objective measure of the evolving of research within a specific field.

On April 15th 2009 a series of queries related to COST 731 topics was submitted to the "ISI Web of Knowledge" database (Table 1). The search was limited to papers within the following subject areas:

- water resources or,
- environmental sciences or,
- geosciences, multidisciplinary or,
- meteorology and atmospheric sciences.

The results of the query are divided into hits up to 2005, roughly the time of writing of the COST 731 proposal, and hits after 2005. It is immediately apparent that the field of hydrological and meteorological modelling and especially the combination of the two has experienced increasing scientific attention in the last few years. Queries for "AND HYDRO* AND MODEL" and for "AND METEO* AND MODEL" show about the same amount of papers for the whole period up to 2005 as from 2005 on. In contrast the hits for "AND HYDRO* AND METEO* AND MODEL" almost double after 2005, this demonstrates a substantial interest in combining meteorological and hydrological models to form model chains. This development probably reflects advances in computing sciences. Powerful computers are able to produce timely forecasts which meet the deadline constraints and extend lead-times for flood warnings, and this despite the increasing complexity of model chains.

In addition there has been an increase of about 75% to 100% in citations of these papers after 2005 compared to the whole period up to 2005 for all queries except the one for "AND METEO* AND MODEL". This reflects the fact that operational experience with determining uncertainty in the meteorological community dates back at least to the early nineties (Palmer and Buizza, 2007).

Surprisingly, despite the huge number of publications which emerged after 2005, only about a quarter of the recent cites refers to recent papers. It seems that the new applications are emerging on the basis of previous work of the groups presenting their own implementation of probabilistic hydrometeorological forecast models. This might also indicate that there is no innovative approach generating a large impact in terms of citations so far.

In addition, in the last few years a significant number of journal special issues dedicated to operational hydrometeorology and flood forecasting have been produced. They reflect the development towards the present state in probabilistic flood forecasting (Table 2).

Table 1
Bibliometric analysis by queries through the "ISI Web of Knowledge" on April 15th 2009.

(Uncertainty or probabilistic or ensemble) and …	First hit	Up to 2005		After 2005		
		Papers	Cites	Recent papers	Recent cites	Cites to recent papers (%)
HYDRO* AND MODEL*	1973	1263	10,828	1324	20,644	25.3
METEO* AND MODEL*	1990	438	5110	495	6354	24.7
HYDRO* AND METEO* AND MODEL*	1991	74	809	133	1451	28.5
FORECAST	1972	684	7085	757	12,643	23.4

Table 2
Selection of recent Special Issues of ISI Journals on topics related to COST 731.

Title	Journal	Vol	#	Topics	Selected articles
HYREX: the HYdrological Radar Experiment	HESS 2000	4(521–679)	12	- Radar precipitation measurements for hydrological purposes - Stochastic space–time rainfall forecasting for real time flow forecasting - Short period forecasting of catchment-scale precipitation - Sensitivity runoff ⟺ rainfall data at different spatial scales	Mellor et al. (2000a,b) Bell and Moore (2000)
Hydrological and meteorological aspects of floods in the Alps	HESS 2003	7(783–948)	11	- Model parameterization - Flood forecasting - Model comparison	Bacchi and Ranzi (2003) Bach et al. (2003) Benoit et al. (2003)
Scientific results from the MAP SOP field experiment.	QJRM 2003	129(341–899)	25	- Orographic precipitation events, Alpine storms - Airflow within/across Alpine river valleys/Alpine ridge - Validation tools for atmospheric models	Ranzi et al. (2003) Jasper and Kaufmann (2003) Reitebuch et al. (2003)
Quantitative Precipitation Forecasting II	JOH 2004	288(1–126)	15	- Rainfall assimilation - Convection-resolving precipitation forecasts - Development of precipitation forecasting and its predictability	Walser and Schär (2004) Orlandi et al. (2004)
VOLTAIRE — an EU framework programme	MetZet 2006	15(5)483–573	10	- Variation of weather radar sensitivity - Radar data quality control - Downscaling model for radar-based precipitation fields - Improvements in weather radar rain rate estimates	Golz et al. (2006) Franco et al. (2006)
Advances in radar, multi-sensor and hydrological modelling methods for flash flood forecasting	NHESS 2006	6,7		- Weather radar beam propagation - Analysis of sever convective events/dual polarisation Doppler radar - Spatio-temporal precipitation error propagation in runoff modelling - Combined clutter and beam blockage correction technique	Berne and Uijlenhoet (2006) Fornasiero et al. (2006)
Uncertainties in hydrological observations	HESS 2006	10(755–601)	9	- Soil moisture from point observations; soil physical data - Remote sensing observation for model calibration - Uncertainties digital elevation models and land use data - Geological and hydrogeological data - Stochastic simulation experiment → radar rainfall uncertainty	Pappenberger et al. (2007) Van der Keur and Iversen (2006) Uijlenhoet and Berne (2008)
MAP findings	QJRM 2007	133(809–1071)	16	- MAP results and findings, benefits, lessons - Quantitative Precipitation Forecasting in the Alps - Hydrological aspects of MAP - Data assimilation - Inter-domain cooperation	Richard et al. (2007) Ranzi et al. (2007)
Hydrological prediction uncertainty	HESS 2007	11, 12, 13	6	- Skill and value of hydrological ensemble predictions - Bias-correction methods, verification tools, uncertainty analysis	Roulin (2007) Xuan et al. (2009)
The German Priority Program Spp1167 "Quantitative Precipitation Forecast"	MetZet 2008	17(6)703–948	17	- Assimilation of radar and satellite data in mesoscale models - Scale dependent analyses of precipitation forecasts and cloud properties - Hybrid convection scheme - Systematic errors in QPF	Milan et al. (2008)
HEPEX Workshop: Stresa, Italy, June 2007	ASL 2008	9(27–102)	11	- HEPEX → aims, challenges, progress - Probabilistic prediction: value, error correction and evaluation of ensembles - MAP D-PHASE: real time demonstration - Probabilistic quantitative Precipitation Forecast for flash flood forecasting - Hydrological aspects of meteorological verification	Buizza (2008) Pappenberger et al. (2008) Zappa et al. (2008) Bogner and Kalas (2008)
Propagation of uncertainty in advanced meteo-hydrological forecast systems	NHESS 2008	8	7	- Uncertainty in radar-based data - Verification of operational Quantitative Discharge Forecast - End-user requirement for surface water runoff design - Model intercomparison	Szturc et al. (2008) Jaun et al. (2008)
Flood forecasting and warning	MA 2009	16(1)	11	- Long lead times - Flood forecasting in England - Radar rainfall nowcasting for flood forecasting and warning	Collier (2009) Werner et al. (2009) Werner and Cranston (2009)

2.2. The structure and objectives of COST 731 Action

COST is an intergovernmental framework for European Cooperation in Science and Technology, allowing the coordination of nationally-funded research on a European level. COST contributes to reducing the fragmentation in European research investments and opens the European Research Area to worldwide cooperation, thus ensuring that Europe holds a strong position in the field of scientific and technical research for peaceful purposes, by increasing European cooperation and interaction in nine key domains, one of which is the Earth System Science and Environmental Management (ESSEM, see www.cost.esf.org). The COST 731 Action was proposed within the ESSEM Domain and launched in mid 2005 for a five-year period. By the end of the Action in 2010 23 countries joined the Action: Australia, Belgium, Cyprus, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Italy, Luxembourg, Netherlands, Norway, Poland, Portugal, Rumania, Spain, Sweden, Switzerland, and United Kingdom.

COST 731 was designed to address the quantification of the uncertainty in meteorological observation and forecasting along with their effect on hydrological forecasting, and the subsequent impact on the decision making process. Dealing with uncertainties in a flood forecasting and warning production chain (Fig. 2) in a consistent way requires the following general stages:

1. atmospheric observation (e.g. precipitation by radar) and quality characterization;
2. assimilation of atmospheric observations into a NWP system;
3. probabilistic atmospheric forecasting in a NWP system (ensembles, neural networks, others);
4. hydrological modelling with atmospheric observations and forecasts, including their associated uncertainties;
5. flood response decision making (especially protection vs. evacuation), management decisions during the event and public warnings.

Radar scientists are mainly concerned with stage 1, NWP modellers with stages 2 and 3. Hydrologists deal with stage 4 but, at present, without making extensive use of radar precipitation estimates and NWP precipitation forecasts. Learning from the COST 717 experience and recognizing the need for an effective interdisciplinary collaboration in order to deal with the propagation of uncertainty from one part of the forecasting/warning system to the next in a coherent way, Working Groups (WGs) are defined on the interfaces between the participating scientific communities in the following way in order to maximise the interactions:

- WG-1: Propagation of uncertainty from observing systems (radars) into NWP (Rossa et al., 2010a);
- WG-2: Propagation of uncertainty from observing systems and NWP into hydrological models (Zappa et al., 2010);
- WG-3: Use of uncertainty in warnings and decision making (Bruen et al., 2010).

A number of interdisciplinary links were implemented to guarantee the transfer of knowledge among the different scientific communities on an appropriate level and to allow for an effective modelling/decision making chain.

The main objective of the Action was to address issues intimately associated with the quality and uncertainty of meteorological observations from remote sensing and other potentially valuable instrumentation, along with their impacts

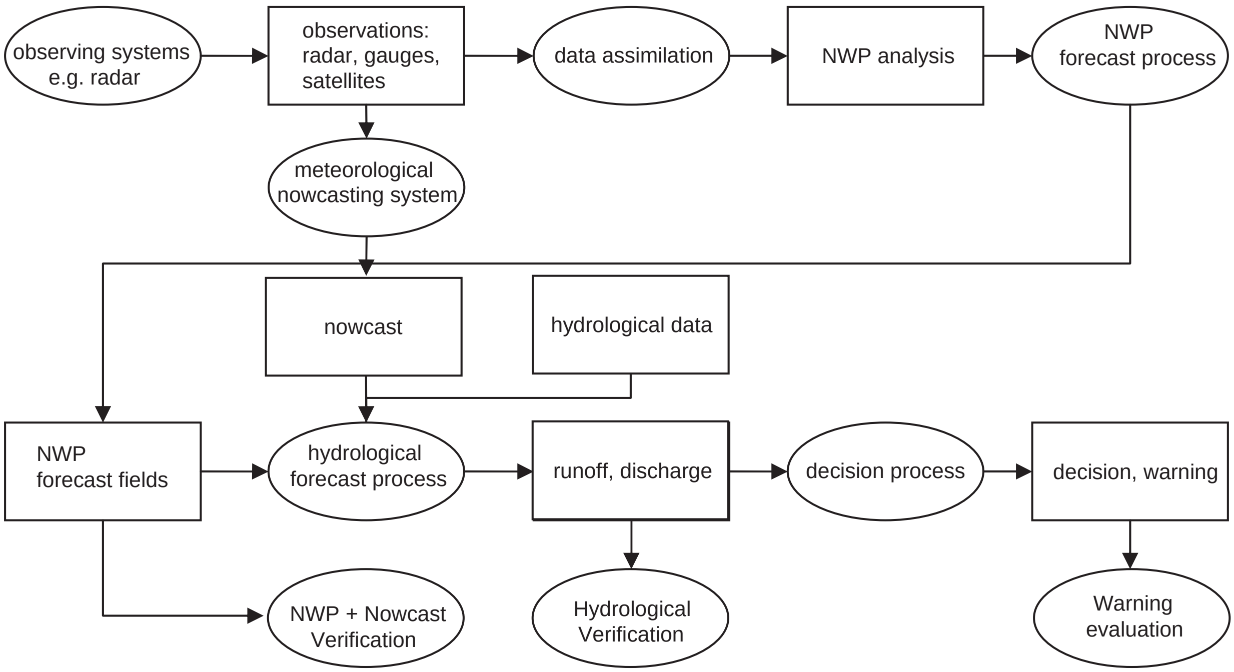

Fig. 2. Schematic to depict the production chain of a flood forecasting, decision making and warning system. Boxes denote products while ellipses denote processes. Note that "NWP" may include any numerical forecast output, deterministic or probabilistic, e.g. ensemble prediction system (EPS) products.

Table 3
Selection on past and ongoing EU framework programmes projects are related to COST 731.

Acronym	Project title	Duration	FP#
TELFLOOD	Forecasting floods in urban areas downstream of steep catchments	1997–1999	FP4
RAPHAEL	Runoff and atmospheric processes for flood hazard forecasting and control http://www.map.meteoswiss.ch/map-doc/NL7/RaphaelProject.htm	1998–2000	FP4
EFFS	European Flood Forecasting System http://effs.wldelft.nl/	2000–2003	FP5
MUSIC	Multi-sensor precipitation measurements integration, calibration and flood forecasting http://www.geomin.unibo.it/hydro/music/	2001–2004	FP5
MANTISSA	Microwave attenuation as a new total improving stormwater supervision administration http://prswww.essex.ac.uk/mantissa/	2001–2004	FP5
CARPE DIEM	Critical Assessment of Available Radar Precipitation Estimation Techniques and Development of Innovative Approaches for Environmental Management http://carpediem.ub.es/	2002–2004	FP5
VOLTAIRE	Validation of multisensors precipitation fields and numerical modelling in Mediterranean test sites http://www.voltaireproject.net/	2002–2006	FP5
FLOODMAN	Near real time flood forecasting, warning and management system based on satellite radar images, hydrological and hydraulic models and in-situ data http://projects.itek.norut.no/floodman/Index.htm	2003–2006	FP5
FLOOD RELIEF	A real-time decision support system integrating hydrological, meteorological and radar technologies http://projects.dhi.dk/floodrelief/index2.asp	2003–2006	FP5
PREVIEW	Prevention, information and early warning pre-operational services to support the management of risks http://www.preview-risk.com	2005–2008	FP6
HYDRATE	Hydrometeorological data resources and technologies for effective flash flood forecasting http://www.hydrate.tesaf.unipd.it/	2007–2009	FP6
FLOODSITE	Integrated Flood Risk Analysis and Management Methodologies http://www.floodsite.net/default.htm	2006–2009	FP6
IMPRINTS	IMproving Preparedness and RIsk maNagemenT for flash floods and debriS flow events http://www.imprints-fp7.eu/	2009–2012	FP7

on hydro-meteorological outputs from advanced forecast systems. This was achieved through specific objectives which can be summarized as follows:

- Radar data assimilation in NWP: provide radar data errors in a form suitable for assimilation schemes, and compare different assimilation techniques for the cloud resolving scale, including nudging, 3- and 4-dimensional variational assimilation and the ensemble Kalman filter techniques and establish their sensitivity to the specification of radar uncertainty.
- Radar data quality description: in collaboration with OPERA (Holleman et al., 2006), update the NWP user requirement for radar data to assist operational data providers.
- Radar ensembles: Investigate methods for generation of ensembles based on uncertainty in radar observations.
- Understand uncertainty: clarify and understand the meaning of uncertainty and establish and agree upon ways to measure and express them.
- Use of uncertainty in hydrological models: establish a standard methodology which has the potential to be a reference in the future, and provide feedback for improvement of meteorological input data (Section 2.4).
- Methodology transfer: explore the potential of techniques used to quantify uncertainty commonly used in meteorology applied to hydrology, and promote them to end users.
- Test beds as proof of concept: set up (a) European test bed(s) in which to run a demonstration project as a proof of concept for probabilistic flood forecasting systems. Test beds integrate observation and forecast uncertainty into a hydrological forecast to provide warning uncertainty. A "simulation package" including a hydrological model and all aspects of decision making can be used for presentation, education and training as well as to perform sensitivity studies.

2.3. MAP D-PHASE and HEPEX

Numerous regular contributors of COST 731 have been and are also involved in MAP D-PHASE and HEPEX, two large initiatives on demonstrating the potential of hydrological ensemble prediction systems. The following section gives a short overview on these two projects.

"MAP D-PHASE" is an acronym for Mesoscale Alpine Program Demonstration of Probabilistic Hydrological and Atmospheric Simulation of flood Events in the Alps (Rotach et al., 2009). The MAP D-PHASE initiative was an important element of the COST 731 Action, right from its initial planning. This WWRP (World Weather Research Programme)-approved Forecast Demonstration Project (FDP) D-PHASE was a follow-on project of the Mesoscale Alpine Programme (MAP, Volkert and Gutermann, 2007) to demonstrate the societal impact of MAP by showcasing the progress achieved in high-resolution and probabilistic numerical weather prediction in complex terrain, along with the consequent benefits for hydrological forecasting.

The heart of D-PHASE was a distributed end-to-end forecasting system geared to Alpine flood events which was set up to demonstrate the state-of-the-art in forecasting precipitation-related high-impact weather. The forecast products from 7 ensemble prediction systems (EPS), 23 deterministic NWP models, and 7 hydrological models during the six-month real-time demonstration phase, were prepared in a harmonized way and accessible through a single, central visualization platform. The probabilistic forecasts based on EPS had a lead time of a few days. Shorter-range forecasts were based on high-resolution atmospheric and hydrologic models for selected regions or catchments. These were complemented with real-time nowcasting and high-resolution observed information. Throughout the forecasting chain, warnings were issued and re-evaluated as the potential flooding event approached, allowing forecasters and end users to issue alerts and make decisions at appropriate times. A first insight into MAP D-PHASE with a focus on operational ensemble hydrological simulations is presented in Zappa et al. (2008), while an account on the end user perspective was compiled by Frick and Hegg (2011-this issue).

The Hydrological Ensemble Prediction Experiment (HEPEX) was launched as a bottom-up process by scientists and users at an ECMWF workshop in 2004. This international research activity is designed to address questions related to end-to-end forecast systems in order to build useful systems and to promote their rapid development and deployment. Schaake et al. (2007b) present some of the key scientific questions associated with the major components of a probabilistic hydrological forecast system, including calibration and downscaling of ensemble weather and climate forecasts, hydrological data assimilation, and user issues. Additional science questions were defined at the third HEPEX workshop held in Stresa in June 2007 (Thielen et al., 2008). Approximately ten site specific test beds, as well as four multidisciplinary test beds have been activated, focusing on one or more clearly defined HEPEX science questions. These have the potential to develop data resources needed for community experiments to address all of the scientific questions, and are expected to include active user participation.

COST 731 joined with both the MAP D-PHASE (in Bologna, May 2008) and HEPEX (in Toulouse, June 2009) communities for common workshops with the goal of sharing expertise and establishing scientific collaboration.

2.4. Link to EU FP projects

A number of recent and ongoing EU projects deal with flood risk management and flood forecasting issues, many of which are related to the objectives of the COST 731 Action. Since the end of the 1980s, the European Commission has been continuously supporting research on floods and other natural hazards (see Table 3 for a list of projects and references to their websites). At the beginning, the EC-sponsored projects were focused on understanding phenomena, identifying concepts and problems, and defining suitable models. As a result of the 3rd Framework Programme (FP) and the 1st phase of the 4th FP, the European Scientific Community had a solid scientific and integrated research infrastructure in this field. Then there was a switch in focus from basic research towards applied research and its problem-solving approach. Stimulated by the knowledge gathered by the RIBAMOD Concerted Action (Balabanis et al., 1999), the EC launched a series of new research projects up to the currently running 7th FP with the declared objective of making substantial progress in the field of flood management. Within the EUROflood research project Parker and Fordham (1996) evaluated flood forecasting, warning and response systems (FFWRS) in the European Union. This research project is a part of the EC funded EPOCH programme (European Programme on Climatology and Natural Hazards). They concluded that despite advances in flood forecasting, FFWRS often under-perform because of unsatisfactory dissemination of warnings and unsatisfactory responses.

Within the 5th FP the European Flood Forecasting System (EFFS, de Roo et al., 2003) aimed to develop a prototype of an European flood forecasting system for lead times of 4–10 days in advance by using the ECMWF NWP ensemble forecast product. The FLOODMAN project developed, demonstrated and validated a prototype information system for cost effective near-real time flood forecasting, warning and management using earth observation data, in particular space borne Synthetic Aperture Radar data, hydrological and hydraulic models and in-situ data. The project FLOOD RELIEF developed and demonstrated on two highly flood prone regional basins (Odra catchment in Poland, Welland and Glen catchments in the UK), a new generation of flood forecasting methods. These basins were selected because they include a wide range of flood producing storms and hydrological regimes over a broad spectrum of spatial and temporal scales. Improved flood forecasting technologies made the results more readily accessible both to flood managers and those threatened by floods. This was achieved by exploiting and integrating different sources of forecast information, including improved hydrological and meteorological model systems and databases, radar, advanced data assimilation procedures and uncertainty estimation, into real-time flood management decision support tools designed to meet the needs of regional flood forecasting authorities.

Within the subsequent EU-FP6 Programme, the FLOODsite Integrated Project (Samuels et al., 2006; Klijn et al., 2008) aimed at an improved understanding of specific flood processes and mechanisms and methodologies for flood risk analysis and management ranging from the high level management of risk at a river-basin, estuary and coastal process cell scale down to the detailed assessment in specific areas. It included specific actions on the hazard of coastal extremes, coastal morphodynamics and flash flood forecasting, as well as understanding of social vulnerability and flood impacts, which are critical to improving the mitigation of flood risk from all causes and thus contributing to FLOODsite's main vision of what a comprehensive approach to flood risk assessment and management ought to encompass. The objectives of the HYDRATE project (Borga et al., 2008; Marchi et al., 2010) included to improve the scientific basis of flash flood forecasting by extending the understanding of past flash flood events, advancing and harmonizing a European-wide innovative flash flood observation strategy, and developing a coherent set of technologies and tools for effective early warning systems. The observation strategy proposed in HYDRATE aimed to collect flash flood data by combining hydrometeorological monitoring and the acquisition of complementary information from post-event surveys. The EU-funded FP6 Integrated Project PREVIEW (PREVention, Information and Early Warning, Mueller et al., 2009) aimed at developing operational geo-information services in support of European civil protection units and local/regional authorities for the management of risks at the European scale. The risk areas covered are flood, fire, windstorm, earthquake, volcanoes, landslide and man-made products and services. Flood forecasting products and services using the ensemble technique are developed, tested, and tailored for flash flood, short-range and medium range fluvial flood forecasting as well as forecasting of so called Northern Floods, triggered by snowmelt.

The European capacity to respond to emergency situations like fire, floods, earthquakes, volcanic eruptions, landslides and humanitarian crisis will be fortified through the SAFER project, a follow on effort of PREVIEW funded by EC FP7.

Among other SAFER intends to integrate the lessons learnt within PREVIEW to consolidate the European Flood Alert System. This will be done by developing a best practice flooding approach based on an enhanced partnership between the public at risk from flooding and the authorities responsible for spatial planning, flood protection and flood emergency response management. IMPRINTS (IMproving Preparedness and RIsk maNagemenT for flash floods and debriS flow events) is part of the ongoing EC FP7 and started in 2009. IMPRINTS aims to contribute to a better preparedness and a better operational risk management for flash floods and debris flow events. These improvements shall help reducing fatalities and economic damage, caused by these kinds of natural hazards. To achieve this, methods and tools are developed which will help risk managers and decision makers in emergency agencies and utility companies to take the necessary measures in good time.

2.5. Considerations on quantifying uncertainty

The Flood Risk and Uncertainty Glossary of the Flood Risk Management Consortium (FRMRC) defines uncertainty as 'a general concept that reflects our lack of sureness about someone or something, ranging from just short of complete sureness to an almost complete lack of conviction about an outcome' (Pappenberger et al., 2006). Probably one of the most common ways for a scientist to view uncertainty is in terms of a mean value which can vary often by plus or minus one, less often by two, and rarely by three units of the variance, typically assuming a Gaussian or Normal probability density function. In complex systems the uncertainties, or errors, can deviate substantially from the Normal distribution to the extent that mean and variance do not have such an intuitive meaning. The Monte Carlo approach is a frequently used approach to exploring the form of the probability density of such complex systems. It simulates a large number of realizations of the system within a given set of constraints to estimate a large number of possible outcomes. If the system, such as a complete hydro-meteorological forecast chain for instance, consists of a number of individual components with a complexity of their own, the Monte Carlo approach can require very significant computing resources, as the total number of realizations can be as large as the product of the realizations necessary for each of the individual components (e.g. Pappenberger et al., 2005).

The sources of uncertainty in a hydro-meteorological forecast system include incomplete observations, approximate forecast models due to unavoidable simplifications or errors in the mathematical representations of the processes, and the chaotic dynamics of the atmosphere, all of which concur to make meteorological, and therefore also hydrological, predictions essentially probabilistic (Palmer, 2001). Wilks (2006) stresses that the randomness of a system does not imply that it is unpredictable or void of information, but rather not precisely predictable. Probability provides the tools to represent the precision of the forecast and this in turn can be valuable information.

Ideally, the prediction of any future state of a hydro-meteorological system should be expressed in terms of a probability density function, which allows making a statement about the system's expected value and the associated spread or uncertainty. In this sense, a good forecast reduces the uncertainty regarding the future state, i.e. the predictive uncertainty which is the 'total uncertainty about, for instance, a hydrological predictand. This is expressed in terms of a probability distribution conditional on all available information and knowledge, where the knowledge typically is embodied in a model (Krzysztofowicz, 1999). In order to know the predictive uncertainty it is necessary to characterize a forecast system with respect to the actual observations which, at the time of the forecast, are unknown.

Todini and Mantovan (2007) strongly suggest distinguishing predictive uncertainty from model uncertainty, implying that only by rigorous Bayesian learning using historical observations can predictive uncertainty effectively be assessed. They point out that too often model uncertainty is evaluated instead, such as in the generalized unbiased linear estimator (GLUE) method for estimating uncertainty (Beven and Binley, 1992). For complex systems, however, this kind of conditioning may not be practicable because there are too many unknowns and the assumption that every residual is informative may not automatically hold (Beven et al., 2008). On the other hand, Beven et al. (2008) retain that if a (good) model is able to span past observations, then it should be expected to also span future observation, in which case the model uncertainties would be an assessment of the predictive uncertainty. Decision makers need to learn to incorporate this uncertainty into the decision making process, while scientists need to decide how to deal with the known and unknown errors of their prediction systems.

To achieve a common understanding of uncertainty it is of utmost importance to develop standardized ways and measures to quantify the uncertainties. Only by agreeing on the various possibilities for quantifying uncertainty and how to interpret them (e.g. skill scores) will it be possible for the different disciplines to work together efficiently and to combine the advantages of the many models and systems in use.

3. Review of recent interest in dealing with uncertainties

To appreciate the challenge of providing a reliable flood forecast it is sufficient to realize that precipitation is the most difficult of the atmospheric parameters to forecast and observe (Sevruk, 1996; Walser et al., 2004). Quantitative precipitation estimation (QPE, Germann et al., 2006) and Quantitative Precipitation Forecasting (QPF, Richard et al., 2007) are key tools for quantifying the potential for flooding, especially on short time scales and for relatively small river and urban catchments. Collier (2007) reviews the factors that limit the predictability in flash flood forecasting and discusses the uncertainties involved in QPE and QPF. Over the last decade, operational meteorological models have achieved spatial scales compatible with operational hydrological models for large, medium, and medium-to-small scale catchments (Volkert and Gutermann, 2007). Meteorological input uncertainty is usually assumed to be the largest source of uncertainty in the prediction of floods, at least for lead times of 2–3 days. Moreover, statistical treatment of any kind in connection with rare events is extremely difficult and again provides little predictive skill (Frei and Schär, 2001). In the 1990s, atmospheric modellers started to use ensemble prediction systems to assess the uncertainty involved in forecasting precipitation in time and space and to gain additional information on the characteristics of possible events (e.g. Palmer and Buizza, 2007).

However, the actual threat to society that potentially occurs only becomes effective through the involvement of the hydrosphere. In other words, after the prediction of a precipitation event with a certain uncertainty, the uncertainty in the prediction of a related potential flooding event must take catchment behavior and anthropogenic behavior into account (e.g. Jaun et al., 2008). Finally, the possible action taken by the appropriate authorities (or groups with a certain economic interest) again is based on or can benefit from knowledge of the uncertainty of the modelling chain that provided the forecasts of potential damage.

Looking at the process from the reverse angle, uncertainty translates into sensitivity: it is as important for the modeller to know how uncertain the input data is to assess the uncertainty of the model result, as it is for the data provider to know how sensitive the model results are to variations and uncertainties in a certain input variable. For example, the relevance of uncertainty in the precipitation field used as input into a hydrological model depends strongly on the size of the catchment. The larger the catchment the stronger it acts to effectively low-pass filter the variations and uncertainties in the precipitation input. For smaller catchments, however, the details in the precipitation field and the corresponding uncertainties may be determining as a peak may or may just not fall into the catchment. Also, not all parameters of a hydrological model will have the same influence on the representation of the flood peak.

3.1. Using imperfect precipitation observations

Meteorological observations have an uncertainty which should be assessed and expressed in a suitable way. Quantitative precipitation estimation (QPE), both from rain gauge networks and meteorological radars, is traditionally expressed in a deterministic way, i.e. without specifying any 'error bars'. Germann et al. (2006) established an error description of radar QPE by comparison with a dense rain gauge network in complex terrain. Recently, several approaches to estimating the uncertainties in radar QPE have been proposed. They are all based on the recognition that the number of error sources limits the accuracy with which radars can measure both reflectivity and Doppler velocities. Such errors are discussed by Joe (1996), Saltikoff et al. (2004), and Michelson et al. (2005) among others. Some significant problems associated with estimating precipitation from radar information include clutter, anomalous beam propagation, attenuation, shielding, and variations of the vertical profile of reflectivity. Doppler wind measurements might be affected by aliasing where velocities higher than the unambiguous velocity are folded back into the fundamental velocity interval. On the other hand, well calibrated radars can, over flat terrain and at ranges shorter than about 100 km, deliver generally good QPE, i.e. within typical uncertainty limits ranging from 1 to 3 dB in terms of QPE (e.g. Germann et al. 2006, Rossa et al. 2010a).

In recent years advanced quality control and characterization schemes for radar data have been developed (e.g. Friedrich et al., 2006; Parent du Châtelet et al. 2006). These schemes are now ready to be applied in NWP and hydrological models. On the other hand, the synergy between radar and NWP model data can lead to improved QPE and improved radar data quality characterization. The increasing attention which is being payed to radar data quality led, for instance, to the WMO project on Radar Quality and Quantitative Precipitation Estimation Intercomparison (RQQI) with the aim of identifying best practices in QPE.

The EUMETNET OPERA programme ("Operational Programme for the Exchange of weather RAdar information", http://www.knmi.nl/opera) provides a European platform for the exchange of expertise on operationally-oriented weather radar issues and where management procedures are optimized in support of applications where radar-based information is required (Michelson et al., 2005). The successful collaboration between OPERA and COST 717 ("Use of radar observations in hydrological and NWP models"; Rossa et al., 2005b) is continued with the COST 731 Action. In OPERA phase 2 (2004–2006) a generalized framework has been developed to facilitate the propagation of uncertainty information at the interface between weather radar and meteorological and hydrological applications (Holleman et al., 2006, 2008). Due to the amount of interest in quality information, this work is continued in OPERA phase 3 (2007–2011).

Several groups within COST 731 have proposed methods of making use of the quality description or error characteristics of the radar QPE to formulate a probabilistic, or ensemble, QPE (e.g. Krajewski and Georgakakos, 1985). The originally retrieved precipitation field is perturbed with a stochastic component, which has the appropriate space–time covariance structure. To determine this error structure Germann et al. (2009) utilize the error climatology of radar QPE assessed by comparison with a rain gauge network, while Schröter et al. (2011-this issue) attempt a real time comparison against a best possible, or benchmark, QPE field. Einfalt et al. (2010) circumvent the challenge of defining the actual precipitation error by using a radar data quality index which is then translated into an error. Pegram et al. (2011-this issue) take yet another, pragmatic approach in that they identify and separate the noise from the signal in a radar QPE field and set the noise equal to the random error, which then is used for ensemble generation. This procedure, however, relies on QPE fields for which systematic errors have been accounted for separately, in particular the biases. Ahrens and Jaun (2007), on the other hand, introduce statistical interpolation of the measurements of a dense rain gauge network to produce gauge based precipitation ensembles. Clark and Slater (2006), Bellerby and Sun (2005), and Bellerby (2007) provide examples of ensemble techniques for rain-gauge and satellite data.

3.2. Using imperfect models

In numerical weather prediction (NWP) modelling there seems not to be a simple way to make assumptions about the error structures of NWP QPF based on observations from rain gauges or radar QPE. This makes it difficult to adapt the statistical process used for radar ensemble QPE (Section 3.1) to determine the corresponding NWP QPF uncertainty. As a matter of fact, the meteorological community tackled this problem from a different angle, i.e. by recognizing that neither the models nor the initial conditions from which they are integrated are perfect. Hereby, the resulting model sensitivities to both specific parameters and the initial conditions is taken as an approximation of the predictive uncertainty (e.g. Palmer et al., 2005). These well known approaches of the meteorological community to produce probabilistic ensembles (Molteni et al., 1996) were taken up by the hydrological community in the last few years (see Cloke and Pappenberger, 2009 for a review).

Insufficient spatial and/or temporal resolution, both in the observations and the modelling, along with a general decrease of predictability when going from larger to smaller scales, can be viewed as another source of uncertainty. NWP models describe larger scale features, such as the advection of a cyclone or an active front for example, quite well even on time scales of a few days, although their smaller-scale details in terms of QPF may be erroneous. When going to the convective scale, models are too approximate, while observations are generally too coarse to represent the convective environment well enough. This source of uncertainty is especially relevant when the resulting QPF is used at smaller scales, e.g. probabilistic global-scale QPF, typically delivered on a grid with a mesh size of tens of kilometers, for a small to medium-scale river catchment with a size of the order of 100–1000 km^2. To deal with this dilemma, a number of statistical downscaling techniques were developed with the aim of generating precipitation ensembles which are consistent with the NWP QPF value but have prescribed statistical properties, which give, most importantly, indications on how intense precipitation can expected to be on any area smaller than the NWP model mesh. These statistical approaches are based on multifractal cascades (Deidda, 2000; Seed, 2003), autoregressive models (Ferraris et al., 2003), and analogue methods (Wetterhall et al., 2005; Gutiérrez et al., 2004). Because there is a significant scale gap between EPS derived from NWPs and typical hydrological models the EPS QPFs must be downscaled or disaggregated for scale correction. Furthermore, the derived ensembles may have to be adjusted for bias or for spread, i.e. representing the appropriate range of uncertainty (Cloke and Pappenberger, 2009).

Convective-scale NWP EPS is a very recent approach to assessing QPF uncertainty at small scales, both in space and time (Gebhardt et al., 2010-this issue). Unlike their medium-range (2–5 days) counterparts for which a sound theory based on the synoptic-scale error growth exists, there does not seem to be a promising analogous procedure based on convective-scale error growth (Hohenegger and Schär 2007). Convective-scale ensembles can be produced by combining forecasts started at different times to form lagged-time ensembles (e.g. Mittermaier 2007), or by perturbing boundary conditions and/or initial conditions coming from a coarser scale ensemble. Recognition of the fact that several processes related to convection are only approximate in the models is used to perturb related model parameters. Keil and Craig (2011) examined forecast uncertainty of a convection-permitting EPS and identified a flow dependence. During weak synoptic-scale forcing perturbations of the model physics are more useful, whereas the lateral boundary conditions are more important when precipitation is forced by the synoptic-scale flow. Another promising approach comes from ensemble Kalman Filter type assimilation schemes, which are being developed in the COSMO consortium to include radar data at the convective scale. These do not only deliver the most probable state of the atmosphere given the available observations, but also quantify uncertainty and thus provide probabilistic information for the initialization of the ensemble forecasts (Hunt et al. 2007).

Input uncertainties are only one of the factors that influence the uncertainties in the output of hydrological models. There are other important sources of uncertainty for instance initialization uncertainty associated with the assumptions made about the initial state of the catchment, the uncertainty of model parameterization and model structure and uncertainties in measurements (Vrugt et al., 2005; Jaun et al., 2008). Ferraris et al. (2002) distinguish between two major sources of uncertainty in the coupled meteo-hydrological forecasting chain. They define an “external” uncertainty associated with numerical approximations, boundary and initial conditions at the meteorological scale and the “internal” uncertainty associated with the hydrological processes involved. In hydrological modelling, the uncertainty is usually conceptually divided into contributions from model structure, model parameters, initial conditions, and meteorological inputs (e.g. Beck, 1987). Model parameters can be divided into physical parameters which can be directly measured, empirical parameters which can be experimentally determined, and conceptual parameters used in approximating process equations which need to be optimized, i.e. indirectly determined. There are a number of methodologies developed for this purpose (see e.g. Mantovan and Todini, 2006), GLUE being one of these. In the past, simulation models have optimized their parameters as if they were steady over time. But in fact representing the time varying nature of hydrological responses related to seasonality and the changing antecedent conditions in the system is an interesting aspect of the problem of finding acceptable models (Choi and Beven, 2007). Two studies addressing this problem (Wagener et al., 2003; Freer et al., 2003) confirm that hydrological processes switch their dynamic behavior between different seasons or periods and this is not expressed properly in most models. Choi and Beven (2007) formulate and evaluate a GLUE-based approach for multi-period and multi-criteria model conditioning of a physically-based distributed model (TOPMODEL) with time-varying hydrological data. In this approach the model calibration is based on identifying periods of different hydrological characteristics. They classified different hydrological periods using the so-called Fuzzy C-means clustering technique. Different parameter sets were determined for each individual cluster. Such multi-period conditioning reduces the forecast uncertainty.

3.3. Methodology transfer from the atmosphere to the hydrosphere

The meteorological community has been using probabilistic forecasting operationally for almost two decades. The most common methodology is Monte Carlo-based ensemble prediction systems (EPS, Palmer and Buizza, 2007). Even if blueprints for ensemble streamflow forecasting are older than one decade (e.g. Schaake et al., 1998; Droegemeier et al., 2000), probabilistic forecasting is still relatively novel in hydrology. However the transfer of the established methodologies used in meteorology may lead to an accelerated development of similar approaches for hydrological purposes. Cloke and Pappenberger

(2009) give a comprehensive review of ensemble techniques. They describe the scientific drivers behind the use of EPS in flood forecasting, they critique some limitations of the case studies in the literature, and highlight some remaining key challenges in using EPS for flood forecasting. The latter include the conceptual complexity of the total uncertainty in the resulting forecast, the limits to improvement posed by the present-day computing resources, and the difficulty of communicating forecast uncertainties adequately.

In this context Xuan et al. (2009) contribute to a better understanding of the implications of the spatial/temporal variability of rainfall forecasts applied in the flood forecasting environment. They used a short-range (24 h) ensemble QPF system to produce rainfall forecasts (2 km weather model grid) to drive a distributed rainfall–runoff model (500 m grid size). On the one hand, they establish the potential value of ensemble forecasts for flood forecasting by concluding that ensemble hydrological forecasts driven by ensemble rainfall forecasts can produce results comparable to forecasts driven by rain-gauge data. On the other hand they also reveal that the bias, especially the common underestimation of rainfall at fine scale, can lead to unrealistic low river flow forecasts.

Verification of such probabilistic forecasts is a particular challenge (Demargne et al., 2009; Cloke and Pappenberger, 2009; Demargne et al., 2010; Brown et al., 2010), in that 'right and wrong' no longer have well-defined meanings when it comes to a single forecast observation pair, or a single case study. Verification needs to take the frequency of occurrence of events into account and requires longer time series. An example is presented by Jaun and Ahrens (2009). Their analysis of two years of hydrological ensemble hindcasts for the upper Rhine catchment shows that the ensemble is able to represent the uncertainty for a variety of different weather situations with an appropriate spread-skill. Roulin (2007) evaluates a hydrological ensemble prediction system using verification methods borrowed from meteorology. To advance adequate procedures to evaluate flood forecasts Cloke and Pappenberger (2008) present a six-step approach for screening new forecast performance measures tailored for use with extreme events in hydrological applications. Some open questions remain on the need for specific components for hydrological applications and how far the "meteorological way of doing" is adequate/applicable for hydrological questions.

Schaake et al. (2007a) describe a technique, used at the US National Weather Service, for generating an ensemble from a single-valued forecast of precipitation and temperature. They divide the spatial domain into subbasins and the time period into model time-steps. They then construct a joint distribution of observations and forecasts and use it to generate ensemble members using the "Shaake shuffle" (Clark et al., 2004). The method demonstrates skill in forecasting both temperature and precipitation for at least five days lead time, but requires a long record of past data for model calibration. The associated estimates of uncertainty are scale dependent (Weygandt et al. 2004).

3.4. Radar data assimilation: one avenue to improving SRNWP QPF

Assimilation of radar data is a major challenge for high-resolution numerical weather prediction models, especially the newest generation of models that explicitly simulate deep convection. Nevertheless, it is a promising avenue for hydrological applications in small river catchments, as it has the potential to bridge the temporal gap between radar-derived QPE and nowcasts and short-range NWP QPF (Collier, 2007). Radar reflectivity measurements are the standard data source for characterizing the spatial distribution of precipitation, and one would expect significant benefits from using this information in the initial conditions of an NWP forecast. However, the nonlinear relationship between reflectivity and the NWP model variables that describe precipitation, the lack of observations that provide a consistent description of the cloud-scale dynamics, and the general low predictability of the atmosphere on small scales, combine to make it difficult to assimilate reflectivity in such a way that the model will retain the information and produce an improved forecast over a longer period of time.

Research on data assimilation methods for models with kilometer-scale resolution (so-called cloud-resolving, or cloud-permitting models) is still in its infancy (Sun, 2005), and many methods are being explored (Sun and Crook, 1998; Caya et al., 2005; Kawabata et al., 2007). The most common method in operational use at the time of writing is latent heat nudging (LHN: Jones and Macpherson, 1997; Leuenberger, 2005; Leuenberger and Rossa, 2007; Montmerle et al., 2007; Stephan et al., 2008; Dixon et al., 2009), although a variety of more advanced techniques have been studied in research contexts. Operational experience suggests that the impact of the assimilated radar data is often short-lived, e.g. a couple of hours, although individual cases can show a much longer lasting positive impact. For example, Stephan et al. (2008) report that within their two month trial period assimilation of radar data on occasions continued to have a strong positive impact after 6 h of free forecast. Dixon et al. (2009), on the other hand, report on five convective cases in which the data assimilation has a dramatic impact on skill during both the assimilation and subsequent forecast periods on nowcasting time scales. Rossa et al. (2010b) confirm this finding for an exceptional case of convection which caused flash flooding and documented the benefit of radar rainfall assimilation in hydrological forecasting. The fact that the impact of the radar data normally decreased rapidly in the first 4 h of the free forecast is likely to be linked to the short life time and predictability of cumulus convection, as well as to deficiencies in current data assimilation methods, particularly those of LHN. Craig et al. (2011) found that the length of time that a high resolution model retains information from assimilation of radar rainfall data is proportional to a convective adjustment time scale. When convection is controlled by the large-scale flow, the convective time scale amounts to few hours and the impact time is short, whereas during weak large-scale forcing situations, when convection is triggered by local-scale phenomena like orography or boundary layer processes, the impact is considerably longer.

3.5. Uncertainty cascading

The determination of uncertainty in a complex forecast system is a formidable task, which involves systematic cascading, or propagation, of the uncertainty of each of the individual system components through the entire system.

Given the often heavily non-linear nature of the modelling system components it is difficult to use traditional linear statistical methods for assessing the overall system uncertainty. Pappenberger et al. (2005) investigated such cascading of model uncertainty from medium range weather forecasts (10-day ahead rainfall forecasts) through the LisFlood rainfall–runoff model down to flood inundation predictions within the European Flood Forecasting System (EFFS). Although there have been a number of studies about uncertainties in real-time flood forecasting which addressed uncertainty cascading through two of these three system components, Pappenberger et al. (2005) were the first to consider the complete modelling chain. Their aim was to assess the uncertainty in the forecast over all combinations of rainfall inputs, runoff forecasts and flood routing models. To reduce the computational demand which arises from applying some form of Monte Carlo technique, they reduced the number of runs required by applying the concept of functional similarity to the parameter sets of the rainfall–runoff model (Pappenberger and Beven, 2004). They found that including medium range rainfall forecasts in the modelling system for real time flood inundation prediction can give useful longer lead times for decision making.

McMillan and Brasington (2008) also studied the end-to-end estimation of uncertainty in a coupled model cascade and produced maps of inundation area estimations at various return periods showing also the uncertainty related with the predictions. Their approach to the problem of computational limitations was different to that of Pappenberger et al. (2005). From the simulated hydrographs they extracted time series of yearly maxima to calculate design flows for various return periods. Design hydrographs for each return period at the 5%, 50% and 95% percentiles of the cumulative distribution were then produced by applying an empirical flow–volume relationship. The design hydrographs were input to the inundation model and transferred the estimated uncertainties on to the inundation estimation map.

In hydrological modelling the diversity of input error sources is an aspect requiring study and needs methodologies to estimate the impact of the different error sources (see Zappa et al., 2011-this issue). The influence of radar rainfall input and model parametric uncertainty on the character of the flow simulation uncertainty in hydrological models has been investigated by Collier (2009) based on a stochastic hydrological model. He compared results derived from the model in the stochastic mode to results from the model in deterministic mode. As rainfall input he used raingauge data, weather radar data (with only ground clutter removed) and a combination of both. He looked at the model performance as a function of the input error in the rainfall and found that the stochastic model produced smaller timing errors than the deterministic model for every type of input error. But the errors in the estimated peak flows were only smaller when the errors associated with the rainfall and runoff input were below the mean input errors used to formulate the stochastic model. Otherwise no advantage was gained from the stochastic model.

Werner and Cranston (2009) compared hydrological forecasts made using predicted rainfall from nowcasts with reference forecasts made with observed rainfall data and with observed radar rainfall in the forecast period. The forecast skills using the predicted rainfall from the radar nowcast were lowest and had highest false alarm rates. But, for fast responding catchments, using the nowcasts was significantly better than not using any rainfall forecast at all. So even if they contain considerable uncertainties the nowcasts contribute positively to the skill of the forecast.

3.6. Test beds and demonstration/training platforms

One of the major quests of the hydro-meteorological community is to have a model chain that integrates all the uncertainties inherent in observations and meteorological forecasts and that is able to propagate these into hydrological forecasts and produce measures of meaningful warning uncertainty. For that purpose test beds are set up for demonstration projects (Schaake et al., 2007b; Zappa et al., 2008) which serve as proof of concept for end users who have never had the opportunity to work with the concept of uncertainty propagation. COST731 WG3 identified and described 7 operational systems with a variety of different objectives (Bruen et al., 2010).

Germann et al. (2009) propose a radar ensemble generator (ensemble of precipitation fields) for use in Alpine catchments using LU decomposition (REAL) which preserves the spatial dependence of the mean and covariance of radar errors. It has been implemented in real-time within the framework of MAP D-PHASE and is linked with the semi-distributed rainfall–runoff model PREVAH (Viviroli et al., 2009) for flash flood modelling in a steep Alpine catchment. It is one of the first experiments of its kind worldwide and can show, operationally, visualizations of ensemble runoff nowcasting with REAL and PREVAH (Fig. 1). Runoff simulation verification on a one year data set for a 44 km^2 subcatchment reveals a reduced bias when using radar QPE or radar ensemble QPE input compared to rain gauge input. Forecast uncertainty measured as scatter between modelled and observed runoff is comparable for radar and rain gauge input, the advantage of the radar ensemble being that it directly provides an estimate of uncertainty for an individual forecast run.

Like MAP D-PHASE, the TIGGE project (THORPEX Interactive Grand Global Ensemble) is a World Weather Research Programme project. It aims to accelerate the improvements in the accuracy of 1-day to 2-week high-impact weather forecasts. Park et al. (2008) describe the preliminary results on comparing and combining ensembles. He et al. (2009) were among the first to use TIGGE for flood warning. They used TIGGE ensemble forecasts to drive a coupled atmospheric–hydrological–hydraulic model system for a mesoscale catchment (4062 km^2) located in the Midlands region of England. They show that the TIGGE database is a promising tool for runoff and inundation forecasting, yielding results comparable to observed discharge and inundation simulations driven by observed discharge.

In 2003, the development of a European Flood Forecasting System (EFAS) was launched as a reaction to the severe Danube and Elbe inundations in 2002. Its goal is to increase the preparedness for floods in trans-national European river basins by supplying medium-range and probabilistic flood forecasting information to local water authorities 3–10 days in advance. The prototype of this system covers all of Europe on a 5 km grid (Thielen et al., 2009).

3.7. Decision making with uncertain information

For some time, hydrologists have recognized the need to integrate uncertain climatic, meteorological and hydrological information into decision making procedures in water resources management (e.g., Georgakakos et al., 2005). In a simplified scheme, four groups of people participate in the overall flood risk management system. Three of them are involved in managing an actual flooding situation: meteorological and hydrological forecasting services, operational water resource managers and civil response managers (Catelli et al., 1998; Morss et al., 2005, 2008). The technical part, i.e. the technological components of the flood forecasting systems used by these flood risk management communities is developed by the meteorological, hydrological, and engineering scientific communities. The application and interpretation of the resulting output involves the fourth group of people, who constitute the social part of the flood risk management system, and must be interpreted on the basis of general information theory as well as social and economic science. These four groups have different needs, perceptions and approaches to handling and using uncertainties. In fact, dealing with uncertainty in hazard warning is necessarily tied to the measure and meaning of the uncertainty information, as well as to how it is communicated and applied. Communication of uncertainty at the interface of science and risk management is not straightforward and needs specific investigation. There are substantial differences between predictive and model uncertainty, the former being relevant for decision makers, while the latter describes the scientifically relevant model-related uncertainty of flood risk assessments (Hall, 2002; Todini, 2004; Beven et al., 2008).

Scientists see scientific uncertainty as a demanding part of the professional domain. On the other hand, managers have to make decisions that often have considerable implications for cost, safety and health of the people and liability of the professionals. So it is not surprising that scientific uncertainty is a rather unwelcome part of decision uncertainty for the professional managers (Faulkner et al., 2007). The communication of risk in flood risk management between scientists and emergency management professionals could be improved by hydrometeorological and engineering models specifically designed to serve as communication tools between the two. In any case, operational flood emergency managers may have difficulties understanding probabilistic or ensemble forecasts if there is no additional explanation or some sort of translation provided (McCarthy et al., 2007).

One important issue to emerge is concern over the use of expected value methods. Experience suggests that its behavioral assumptions are often violated (Machina, 1987) particularly in relation to high-impact, low-probability hazards, such as extreme floods. This has been studied by economists in relation to financial crises (Bussiere and Fratzscher, 2008) and Climate Change mitigation policy (Lange, 2003). The latter describes a combination of the expected utility criterion with a max-i-min approach giving more importance to more extreme events. Birnbaum (2008) discusses a range of alternatives to expected utility, including prospect theory. Others have studied how best to represent risk-aversion (LiCalzi and Sorato, 2006) while Geiger (2000) also studied low probability, high-impact risks. The value of public participation, particularly in comparison with technical experts, in decision making is illustrated by Gamper and Turcanu (2009).

Uncertainty estimates of decision variables, i.e. quantities whose values are set by a risk manager or policy maker, may be viewed as important only to the extent that they contribute to good-decision making (Cox, 1999). In addition to the different kinds of uncertainty, Buizza et al. (2007) propose to consider the so called functional quality of forecasting products. In contrast to the technical quality, another term for model oriented uncertainty, the functional quality of the forecasting products and services has to be taken into account by judging their benefits for decision making. The framework developed for assessing the functional quality encompass the four attributes ‘availability and means of distribution’, ‘content and format’, support, maintenance and training’ and ‘communication of product's technical quality’.

Demeritt et al. (2007) tested how hydrological forecasters involved in real-time flood forecasting handle the uncertainties and possible benefits inherent in EPS. Groups of forecasters were asked to complete a couple of simulated forecasting exercises based on real events. The study showed that the forecasters used the EPS more as a confirmation of their own deterministic models than as a precautionary tool. In situations where the deterministic and the EPS model deviated, they tended to be very cautious in issuing early flood warnings and rather waited and saw how things proceeded. This reaction was explained by the forecasters' consciousness of the possible costs of a false alarm and the associated loss of credibility. So there is a public sensitivity to false alarms in flood forecasting that contrasts with a relative tolerance in weather forecasting. This fundamentally different attitude between meteorologist and flood forecasters in dealing with EPS information is supported by Doswell (2004) who reported that in meteorological weather forecasting not forecasting an event that happens (false negative/miss) is considered worse than forecasting an event that does not happen (false positive/false alarm).

Another experiment by Joslyn and Nichols (2009) investigated how the public handles and understands uncertainty in weather forecasts. The objective question was if uncertainty expressed as frequency is better understood by public than uncertainty expressed as probability. People understood the forecasts better when they were expressed in a probability format, moreover some additional information like a reference class did not improve their understanding.

Georgakakos et al. (2005), in describing the practical use of meteorological–hydrological forecasts in multi-objective reservoir operation, stress the importance of providing demonstration platforms. They list the essential ingredients for a system to be widely accepted as (i) a reliable and adaptive numerical forecast capability, (ii) mutually agreed performance criteria, (iii) a baseline system representing current practice is also available for comparison with the new forecast system, (iv) rigorous, quantitative, intercomparison of methods using historic or real-time data and (v) active and continuing participation of decision-making end-users in workshops. Georgakakos (2004) describes one such system implemented in the Nile catchment.

There is a growing range of powerful quantitative uncertainty, sensitivity, risk and decision analysis techniques, on which flood risk management has begun to draw, but unfortunately, a number of factors conspire to limit the rate and the extent of their uptake. Harvey et al. (2008) discuss these factors and have developed a prototype of a software system named REFRAME to support flood risk analysis. This implements an idealised but realistic flood risk analysis. Another example of a framework of flood risk assessment addressing and taking into account the different sources of uncertainty is given by Apel et al. (2004). Although these tools focus on long term flood risk planning issues they contain most of the ingredients (e.g. tools for assessing the managers perspective of losses by well defined damage functions, inventory and costs of the reaction potential of a certain area) needed for flood forecasting products and formulating guidance material (e.g. Ntelekos et al., 2006).

A consistent result from workshops with end-users in both COST Actions (717, 731) and EU research projects (e.g. CARPE DIEM) is a desire for uncertainty information but a strong call for training in the use of such information. End-users can see its relevance, but are uncertain about how it can best be used. The spaghetti plots often used to represent ensemble results are acknowledged to have some value, but while the producers of these plots worry about undue (in their view) emphasis being placed on the worst member of the ensemble, decision makers do want to know what the worst case scenario entails as well as a history of predictions and outcomes.

4. Outcomes of COST 731

COST 731 can be seen as a timely European initiative to make concerted progress in the field of probabilistic flood forecasting with a particular emphasis on operational applications. The most significant emerging results and trends can be summarized as follows:

- One of the most innovative developments emerging from the COST 731 Action is related to probabilistic quantitative precipitation estimation (QPE) from radar (see review in Section 3). Three groups implemented slightly differing methodologies based on a quality description of the precipitation estimates (Germann et al., 2009; Rossa et al., 2010a, Zappa et al. 2010), while Pegram et al. (2011-this issue) implemented a signal-theory based approach. It is to be seen as a sign of good progress that all of these probabilistic QPE methods are being used in combination with hydrological models for simulation of small river catchments in Switzerland, Spain, and Poland, respectively.
- An increasing number of hydrological models are now using EPS QPF for operational medium- to long-range forecasts for river flow forecasting and water management purposes (Cloke and Pappenberger, 2009; Zappa et al., 2010; Bruen et al., 2010).
- A large number of test beds have been implemented in quasi operational mode, especially during the MAP D-PHASE Operations Phase (Rotach et al. 2009, see Section 2.2), some of them have been online for the duration of MAP D-PHASE in 2007 only, some systems are still providing results in real time.
- A recommendation to OPERA has been made to include a systematic radar data description for European radar data exchange (Rossa et al., 2010a).
- Progress has been made on the convective-scale NWP by radar data assimilation. This is particularly relevant for flash flood prediction in small river catchments where extending the warning lead time is crucial (Rossa et al., 2010a, Rossa et al., 2010b).
- A set of demonstration platforms and tools for communicating uncertainty have been identified and presented to various sets of end users (Bruen et al., 2010). Hereby, visualization of and access to typically very large volumes of data emerged as crucial for the efficient use.

Following on from work started in COST 717 (Rossa et al., 2005b) an earlier, systematic radar data quality description has been taken to the next level. Here, several aspects are worth mentioning in the context of a more extensive and quantitative use of radars, both in NWP and hydrological modelling. The European radar panorama is extremely heterogeneous in terms of type and age of the radar hardware, as well as the observation strategy. Therefore, and given the numerous factors impacting on the measurement, radar data quality is expected to vary significantly from country to country. The EUMETNET Programme OPERA (Holleman et al., 2006) took the COST 731 recommendation to include data quality description in the definitions of the international radar data transmission protocols. As described earlier, an adequate data description is the basis for the generation of probabilistic radar QPE, while statistical data assimilation systems in NWP, like 3/4DVAR, need the observation error to be quantified in a proper form. Finally, hydrological models can take account of radar QPE uncertainty for example in form of QPE ensembles.

A very successful aspect of the MAP D-PHASE is the large number of participating modelling groups, totaling 7 meteorological ensemble prediction systems, 23 deterministic meteorological model, 7 coupled hydrological models covering 43 catchments in the Alpine area (Rotach et al., 2009). The single most important factor of success for this complex initiative was probably the interoperability of all the models: common formats, common warning levels, and common routines to actually determine the warnings from the model outputs rendered the results comparable and therefore highly valuable (Rotach et al., 2009).

From the operational institutions represented in COST 731 there are more than five operational or quasi-operational implementations of integrated flood forecasting systems (Bruen et al., 2010). A particularly interesting example of a test bed implementation is related to the construction of the Zurich railway station underground extension that involved closing 2 of 5 gates through which the river Sihl flows under the railway station (Romang et al., 2011). A hydrometeorological forecasting service started in mid 2008 to run for three years and deliver medium-range probabilistic forecasts of the flow of river Sihl to support the planning of construction work in the river. In critical conditions, the closed gates must be opened, which necessitates stopping construction in the river bed and evacuating the site, and the construction time table is delayed. The loss for not opening the gates, if a flood comes, is very high as a part of the Zurich downtown area would be flooded. This makes an excellent self-contained evaluation exercise and is a unique opportunity available to COST 731.

The Demonstration activities allocated within COST 731 produced a list of existing or potential demonstration platforms or published case studies (including simulation exercises) that could be useful in training or research in the use of uncertainty information. These were discussed at a COST 731 end user workshop in Dublin and include inputs from a wide range of European countries:

- The MAP-D-PHASE visualization platform, covering the greater European Alpine area (Rotach et al., 2009);
- The hydrological ensemble prediction System for Zurich Railway station in Switzerland (Romang et al., 2011);
- The river simulator used by the Swedish HydroPower industry;
- SMHI's WebHyPro system in Sweden (Arheimer et al. 2011-this issue);
- Flood risk management and flood forecasting in the River Rhine basin operated in Germany and Switzerland;
- Flood risk management and flood forecasting in the River Danube Basin;
- The Results of the EU-FP-6 Project PREVIEW;
- The Finnish Flood forecasting and warning system (Vehvilainen et al., 2005).

5. Outlook and future efforts

The potential value of improved flood forecasting capabilities is beyond controversy. This review testifies to the great effort which is being invested in this field in Europe and elsewhere, both by the research as well as by the operational community. The fact that forecasts of this kind are inherently uncertain, a characteristic that will not change even in the future, seems to be increasingly appreciated, as is the need to adequately quantify and formulate this uncertainty and to make proper use of this information in a decision making context. The COST 731 Action 'Propagation of Uncertainty in Advanced Meteo-Hydrological Forecast Systems' is an expression of and contributes to this trend. A particularly positive aspect hereby is that the meteorological and hydrological community, traditionally quite separate, have increased their cooperation in a very significant way.

Avenues of improvement of flood forecasting include the respective improvement of the individual system components, as well as establishing improved combined systems and promote the interpretation of the system output, notably:

- improving radar QPE for small- to medium-scale river catchments;
- improving short-range NWP QPF by making better use of radar and other non-conventional meteorological information, especially at the convection scale;
- improving observations and use of snow cover and soil moisture, both in meteorological and hydrological models;
- extending limited area EPS to forecast ranges of 7–8 days for water management;
- increasing spatial resolution of NWP EPSs, e.g. at convection scale with radar precipitation and wind assimilation;
- implementing and extending to wider areas existing test bed implementations, e.g. to cover the entire Alpine range;
- enhancing end user and decision maker involvement and training in using probabilistic forecast systems;
- establishing Economic-Value Issues (Cost/Loss Analyses; Roulin, 2007) as a tool for tailored decision making.

The COST 731 has ended in mid 2010 but the work will continue on, especially in the scientific networks that have formed as a result of the Action.

Acknowledgements

The numerous writers of COST 731 Memorandum of Understanding are greatly acknowledged for their contribution. All the Management Committee members are thanked for the great inputs to the Action. Carine Petit, Lucia Forzi, and the COST Office staff are especially thanked for their support in the management of the Action. Urs Germann and Massimiliano Zappa profited from funding from the MeteoSwiss Swiss State Secretariat for Education and Research SER (C05.0105). Katharina Liechti is financed through the EU FP7 Project IMPRINTS (Grant agreement no.: 226555/FP7-ENV-2008-1-226555). Michael Bruen's radar flood forecasting work is supported by Science Foundation Ireland (04-RFP-ENMF292).

References

Ahrens, B., Jaun, S., 2007. On evaluation of ensemble precipitation forecasts with observation-based ensembles. Advances in Geosciences 10, 139–144.
Apel, H., Thieken, A.H., Merz, B., Blöschl, G., 2004. Flood risk assessment and associated uncertainty. Natural Hazards and Earth System Sciences 4 (2), 295–308.
Arheimer, B., Lindström, G., Olsson, J., 2011. A systematic review of sensitivities in the Swedish flood-forecasting system. Atmospheric Research 100, 275–284 (this issue).
Bacchi, B., Ranzi, R., 2003. Hydrological and meteorological aspects of floods in the Alps: an overview. Hydrology and Earth System Sciences 7 (6), 784–798.
Bach, H., Braun, M., Lampart, G., Mauser, W., 2003. Use of remote sensing for hydrological parameterisation of Alpine catchments. Hydrology and Earth System Sciences 7 (6), 862–876.
Balabanis, P., Bronstert, A., Samuels, P., 1999. River Basin Modelling Management and Flood Mitigation —Concerted Action. Proceedings of the final workshop, Walllingford. 26 and 27 February.
Barredo, J.I., 2007. Major flood disasters in Europe: 1950–2005. Natural Hazards 42 (1), 125–148.
Beck, M.B., 1987. Water-quality modeling — a review of the analysis of uncertainty. Water Resource Research 257, 1393–1442.
Bell, V.A., Moore, R.J., 2000. Short period forecasting of catchment-scale precipitation. Part II: a water-balance storm model for short-term rainfall and flood forecasting. Hydrology and Earth System Sciences 4 (4), 635–651.
Bellerby, T.J., 2007. Satellite rainfall uncertainty estimation using an artificial neural network. Journal of Hydrometeorology 8 (6), 1397–1412.
Bellerby, T.J., Sun, J., 2005. Probabilistic and ensemble representations of the uncertainty in an IR/Microwave Satellite Precipitation Product. Journal of Hydrometeorology 6 (6), 1032–1044.
Benoit, R., Kouwen, N., Yu, W., Chamberland, S., Pellerin, P., 2003. Hydrometeorological aspects of the real-time ultrafinescale forecast support during the special observing period of the MAP. Hydrology and Earth System Sciences 7 (6), 877–889.
Berne, A., Uijlenhoet, R., 2006. Quantitative analysis of X-band weather radar attenuation correction accuracy. Natural Hazards and Earth System Sciences 6 (3), 419–425.
Beven, K.J., Binley, A.M., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes 6, 279–298.
Beven, K.J., Smith, P.J., Freer, J.E., 2008. So just why would a modeller choose to be incoherent? Journal of Hydrology 354 (1–4), 15–32.
Birnbaum, M.H., 2008. New paradoxes of risky decision making. Psychological Review 115 (2), 463–501.
Bogner, K., Kalas, M., 2008. Error-correction methods and evaluation of an ensemble based hydrological forecasting system for the Upper Danube catchment. Atmospheric Science Letters 9 (2), 95–102.
Borga, M., Gaume, E., Creutin, J.D., Marchi, L., 2008. Surveying flash floods: gauging the ungauged extremes. Hydrological Processes 22 (18), 3883–3885.
Brown, J.D., Demargne, J., Seo, D.-J., Liu, Y., 2010. The Ensemble Verification System (EVS): a software tool for verifying ensemble forecasts of

hydrometeorological and hydrologic variables at discrete locations. Environmental Modelling & Software 25 (7), 854–872.
Bruen, M., Krahe, P., Zappa, M., Olsson, J., Vehvilainen, B., Kok, K., Daamen, K., 2010. Visualizing flood forecasting uncertainty: some current European EPS platforms-COST731 working group 3. Atmospheric Science Letters 11 (2), 92–99.
Buizza, R., 2008. The value of probabilistic prediction. Atmospheric Science Letters 9 (2), 36–42.
Buizza, R., Asensio, H., Balint, G., Bartholmes, J., Bliefernicht, J., Bogner, K., Chavaux, F., de Roo, A., Donnadille, J., Ducrocq, V., Edlund, C., Kotroni, V., Krahe, P., Kunz, M., Lacire, K., Lelay, M., Marsigli, C., Milelli, M., Montani, A., Pappenberger, F., Rabuffetti, D., Ramos, M.H., Ritter, B., Schipper, J.W., Steiner, P., Thielen-Del Pozzo, J., Vincendon, B., 2007. EURORISK/PREVIEW report on the technical quality, functional quality and forecast value of meteorological and hydrological forecasts. ECMWF Tech. Memorandum 516, 1–63 ECMWF Research Dep., Reading.
Bussiere, M., Fratzscher, M., 2008. Low probability, high impact: policy making and extreme events. Journal of Policy Modeling 30 (1), 111–121.
Catelli, C., Pani, G., Todini, E., 1998. FLOODSS: FLOod Operational Decision Support System. In: Balabanis, P., Bronstert, A., Casale, R., Samuels, P. (Eds.), River Basin Modelling Management and Flood Mitigation — Concerted Action. Proceedings of the final workshop, Walllingford, pp. 381–394. 26 an 27 February.
Caya, A., Sun, J., Snyder, C., 2005. A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation. Monthly Weather Review 133 (11), 3081–3094.
Choi, H.T., Beven, K., 2007. Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework. Journal of Hydrology 332 (3–4), 316–336.
Clark, M.P., Slater, A.G., 2006. Probabilistic quantitative precipitation estimation in complex terrain. Journal of Hydrometeorology 7 (1), 3–22.
Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., Wilby, R., 2004. The Schaake shuffle: a method for reconstructing space–time variability in forecasted precipitation and temperature fields. Journal of Hydrometeorology 5 (1), 243–262.
Cloke, H.L., Pappenberger, F., 2008. Evaluating forecasts of extreme events for hydrological applications: an approach for screening unfamiliar performance measures. Meteorological Applications 15 (1), 181–197.
Cloke, H.L., Pappenberger, F., 2009. Ensemble flood forecasting: a review. Journal of Hydrology 375 (3–4), 613–626.
Collier, C.G., 2007. Flash flood forecasting: what are the limits of predictability? Quarterly Journal of the Royal Meteorological Society 133 (622), 3–23.
Collier, C.G., 2009. On the propagation of uncertainty in weather radar estimates of rainfall through hydrological models. Meteorological Applications 16 (1), 35–40.
Cox, L.A., 1999. Internal dose, uncertainty analysis, and complexity of risk models. Environment International 25 (6–7), 841–852.
Craig, G.C., Keil, C. and Leuenberger, D., 2011. Constraints on the Impact of Radar Rainfall Data Assimilation on Forecasts of Cumulus Convection. Quarterly Journal of the Royal Meteorological Society, in revision.
de Roo, A.P.J., Gouweleeuw, B., Thielen, J., Bartholmes, J., Bongioannini-Cerlini, P., Todini, E., Bates, P.D., Horritt, M., Hunter, N., Beven, K., Pappenberger, F., Heise, E., Rivin, G., Hils, M., Hollingsworth, A., Holst, B., Kwadijk, J., Reggiani, P., Van Dijk, M., Sattler, K., Sprokkereef, E., 2003. Development of a European flood forecasting system. International Journal of River Basin Management 1 (1), 49–59.
Deidda, R., 2000. Rainfall downscaling in a space–time multifractal framework. Water Resources Research 36 (7), 1779–1794.
Demargne, J., Mullusky, M., Werner, K., Adams, T., Lindsey, S., Schwein, N., Marosi, W., Welles, E., 2009. Application of forecast verification science to operational river forecasting in the US National Weather Service. Bulletin of the American Meteorological Society 90 (6), 779.
Demargne, J., Brown, J., Liu, Y.Q., Seo, D.J., Wu, L.M., Toth, Z., Zhu, Y.J., 2010. Diagnostic verification of hydrometeorological and hydrologic ensembles. Atmospheric Science Letters 11 (2), 114–122.
Demeritt, D., Cloke, H., Pappenberger, F., Thielen, J., Bartholmes, J., Ramos, M.-H., 2007. Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting. Environmental Hazards 7 (2), 115–127.
Dixon, M., Li, Z., Lean, H., Roberts, N., Ballard, S., 2009. Impact of data assimilation on forecasting convection over the United Kingdom using a high-resolution version of the met office unified model. Monthly Weather Review 137 (5), 1562–1584.
Doswell, C.A., 2004. Weather forecasting by humans — heuristics and decision making. Weather and Forecasting 19 (6), 1115–1126.
Droegemeier, K.K., Smith, J.D., Businger, S., Doswell, C., Doyle, J., Duffy, C., Foufoula-Georgiou, E., Graziano, T., James, L.D., Krajewski, V., LeMone, M., Lettenmaier, D., Mass, C., Pielke, R., Ray, P., Rutledge, S., Schaake, J., Zipser, E., 2000. Hydrological aspects of weather prediction and flood warnings: Report of the Ninth Prospectus Development Team of the US Weather Research Program. Bulletin of the American Meteorological Society 81 (11), 2665–2680.
Einfalt, T., Szturc, J., Osrodka, K., 2010. The quality index for radar precipitation data: a tower of Babel? Atmospheric Science Letters 11 (2), 139–144.
Faulkner, H., Parker, D., Green, C., Beven, K., 2007. Developing a translational discourse to communicate uncertainty in flood risk between science and the practitioner. AMBIO: A Journal of the Human Environment 36 (8), 692–704.
Ferraris, L., Rudari, R., Siccardi, F., 2002. The uncertainty in the prediction of flash floods in the Northern Mediterranean Environment. Journal of Hydrometeorology 3 (6), 714–727.
Ferraris, L., Gabellani, S., Rebora, N., Provenzale, A., 2003. A comparison of stochastic models for spatial rainfall downscaling. Water Resources Research 39 (12), 1368. doi:10.1029/2003WR002504.
Fornasiero, A., Bech, J., Alberoni, P.P., 2006. Enhanced radar precipitation estimates using a combined clutter and beam blockage correction technique. Natural Hazards and Earth System Sciences 6 (5), 697–710.
Franco, M., Sanchez-Diezma, R., Sempere-Torres, D., 2006. Improvements in weather radar rain rate estimates using a method for identifying the vertical profile of reflectivity from volume radar scans. Meteorologische Zeitschrift 15 (5), 483–573.
Freer, J.E., Beven, K.J., Peters, N.E., 2003. Multivariate seasonal period model rejection within the generalised likelihood uncertainty estimation procedure. In: Duan, Q., Gupta, H., Rousseau, S.S.A.N., Turcotte, R. (Eds.), Calibration of Watershed Models. AGU Water Science and Application Series, Washington, p. 346.
Frei, C., Schär, C., 2001. Detection probability of trends in rare events: theory and application to heavy precipitation in the Alpine region. Journal of Climate 14 (7), 1568–1584.
Frick, J., Hegg, C., 2011. Can end-users' flood management decision making be improved by information about forecast uncertainty? Atmospheric Research 100, 296–303 (this issue).
Friedrich, K., Hagen, M., Einfalt, T., 2006. A quality control concept for radar reflectivity, polarimetric parameters, and Doppler velocity. Journal of Atmospheric and Oceanic Technology 23 (7), 865–887.
Gamper, C.D., Turcanu, C., 2009. Can public participation help managing risks from natural hazards? Safety Science 47 (4), 522–528.
Gebhardt, C., Theis, S.E., Paulat, M., Bouallègue, Z.B., 2011. Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries Atmospheric Research 100, 168–177 (this issue).
Geiger, G., 2000. A dynamic account of rational decision making under uncertainty: the case of risk assessment in hazardous technological systems. In: Matties, M., Malchow, H., Kriz, J. (Eds.), International Conference on Systems Science — Integrative Approaches to Natural and Social Dynamics. Osnabruck, Germany, pp. 305–318.
Georgakakos, A.P., 2004. Decision support systems for integrated water resources management with an application to the Nile Basin. In: Elsevier (Ed.), International Federation for Automatic Control Workshop on Modeling and Control for Participatory Planning and Managing Water Systems. Venice, Italy.
Georgakakos, K.P., Carpenter, T.M., Yao, H., Graham, N.E., Georgakakos, A.P., 2005. Integrating climate-hydrology forecasts and multi-objective reservoir management for Northern California. Eos Trans. AGU 86 (12).
Germann, U., Galli, G., Boscacci, M., Bolliger, M., 2006. Radar precipitation measurement in a mountainous region. Quarterly Journal of the Royal Meteorological Society 132 (618), 1669–1692.
Germann, U., Berenguer, M., Sempere-Torres, D., Zappa, M., 2009. REAL — ensemble radar precipitation estimation for hydrology in a mountainous region. Quarterly Journal of the Royal Meteorological Society 135 (639), 445–456.
Golz, C., Einfalt, T., Galli, G., 2006. Radar data quality control methods in VOLTAIRE. Meteorologische Zeitschrift 15 (5), 497–504.
Gutiérrez, J.M., Cofino, A.S., Cano, R., Rodriguez, M.A., 2004. Clustering methods for statistical downscaling in short-range weather forecasts. Monthly Weather Review 132 (9), 2169–2183.
Hall, J.W., 2002. A contingency approach to choice. Civil Engineering and Environmental Systems 19 (2), 87–118.
Harvey, H., Pappé, R., Hall, J., 2008. Reframe: a software system supporting flood risk analysis. International Journal of River Basin Management 6 (2), 163–174.
He, Y., Wetterhall, F., Cloke, H.L., Pappenberger, F., Wilson, M., Freer, J., McGregor, G., 2009. Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions. Meteorological Applications 16 (1), 91–101.
Hohenegger, C., Schär, C., 2007. Atmospheric predictability at synoptic versus cloud-resolving scales. Bulletin of the American Meteorological Society 88 (11), 1783–1793.
Holleman, I., Michelson, D., Galli, G., Germann, U. and Peura, M., 2006. Quality information for radars and radar data. EUMETNET OPERA document, available at http://www.knmi.nl/opera.

Holleman, I., Delobbe, L., Zgonc, A., 2008. Update on the European Weather Radar Network (OPERA), ERAD Publication Series, available at http://erad2008.fmi.fi/proceedings.
Hunt, B.R., Kostelich, E.J., Szunyogh, I., 2007. Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. Physica D-Nonlinear Phenomena 230 (1–2), 112–126.
Jasper, K., Kaufmann, P., 2003. Coupled runoff simulations as validation tools for atmospheric models at the regional scale. Quarterly Journal of the Royal Meteorological Society 129 (588), 673–692.
Jaun, S., Ahrens, B., 2009. Evaluation of a probabilistic hydrometeorological forecast system. Hydrology and Earth System Sciences Discussions 6 (2), 1843–1877.
Jaun, S., Ahrens, B., Walser, A., Ewen, T., Schär, C., 2008. A probabilistic view on the August 2005 floods in the upper Rhine catchment. Natural Hazards and Earth System Sciences 8 (2), 281–291.
Joe, P., 1996. Precipitation at the ground: radar techniques. In: Raschke, E. (Ed.), Radiation and Water in the Climate System. Springer NATO ASI Series, pp. 277–321.
Jones, C.D., Macpherson, B., 1997. A latent heat nudging scheme for the assimilation of precipitation data into an operational mesoscale model. Meteorological Applications 4 (03), 269–277.
Joslyn, S.L., Nichols, R.M., 2009. Probability or frequency? Expressing forecast uncertainty in public weather forecasts. Meteorological Applications 16 (3), 309–314.
Kawabata, T., Seko, H., Saito, K., Kuroda, T., Tamiya, K., Tsuyuki, T., Honda, Y., Wakazuki, Y., 2007. An assimilation and forecasting experiment of the Nerima heavy rainfall with a cloud-resolving nonhydrostatic 4-dimensional variational data assimilation system. Journal of the Meteorological Society of Japan 85 (3), 255–276.
Keil, C. and Craig, G.C., 2011. Regime-dependent forecast uncertainty of convective precipitation. Meteorol. Z., accepted.
Klijn, F., Samuels, P., Van Os, A., 2008. "Towards flood risk management in the EU: State of affairs with examples from various European countries". International Journal of River Basin Management 6 (4), 307–321.
Krajewski, W.F., Georgakakos, K.P., 1985. Synthesis of radar rainfall data. Water Resources Research 21 (5), 764–768.
Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. Water Resources Research 35, 2739–2750.
Lange, A., 2003. Climate change and the irreversibility effect combining expected utility and MaxiMin. Environmental & Resource Economics 25 (4), 417–434.
Leuenberger, D., 2005. High-resolution Radar Rainfall Assimilation: Exploratory Studies with Latent Heat Nudging. ETH, Zürich. 103 pp.
Leuenberger, D., Rossa, A., 2007. Revisiting the latent heat nudging scheme for the rainfall assimilation of a simulated convective storm. Meteorology and Atmospheric Physics 98 (3–4), 195–215.
LiCalzi, M., Sorato, A., 2006. The Pearson system of utility functions. European Journal of Operational Research 172 (2), 560–573.
Machina, M.J., 1987. Decision-making in the presence of risk. Science 236 (4801), 537–543.
Mantovan, P., Todini, E., 2006. Hydrological forecasting uncertainty assessment: incoherence of the GLUE methodology. Journal of Hydrology 330 (1–2), 368–381.
Marchi, L., Borga, M., Preciso, E., Gaume, E., 2010. Characterisation of selected extreme flash floods in Europe and implications for flood risk management. Journal of Hydrology 394 (1–2), 118–133.
McCarthy, S., Tunstall, S., Parker, D., Faulkner, H., Howe, J., 2007. Risk communication in emergency response to a simulated extreme flood. Environmental Hazards 7 (3), 179–192.
McMillan, H.K., Brasington, J., 2008. End-to-end flood risk assessment: a coupled model cascade with uncertainty estimation. Water Resources Research 44 (3).
Meischner, P., Collier, C., Illingworth, A., Joss, J., Randeu, W., 1997. Advanced weather radar systems in Europe: The COST 75 Action. Bulletin of the American Meteorological Society 78 (7), 1411–1430.
Mellor, D., Sheffield, J., O'Connell, P.E., Metcalfe, A.V., 2000a. A stochastic space–time rainfall forecasting system for real time flow forecasting I: development of MTB conditional rainfall scenario generator. Hydrology and Earth System Sciences 4 (4), 603–615.
Mellor, D., Sheffield, J., O'Connell, P.E., Metcalfe, A.V., 2000b. A stochastic space–time rainfall forecasting system for real time flow forecasting II: application of SHETRAN and ARNO rainfall runoff models to the Brue catchment. Hydrology and Earth System Sciences 4 (4), 617–626.
Michelson, D., Einfalt, T., Holleman, I., Gjertsen, U., Friedrich, K., Haase, G., Lindskog, M., Jurczyk, A., 2005. Weather Radar Data Quality in Europe: Quality Control and Characterization. COST 717 document, Brussels, p. 87.
Milan, M., Venema, V., Schuttemeyer, D., Simmer, C., 2008. Assimilation of radar and satellite data in mesoscale models: a physical initialization scheme. Meteorologische Zeitschrift 17 (6), 887–902.
Mitchell, J.K., 2003. European river floods in a changing world. Risk Anal 23, 567–574.
Mittermaier, M.P., 2007. Improving short-range high-resolution model precipitation forecast skill using time-lagged ensembles. Quarterly Journal of the Royal Meteorological Society 133 (627), 1487–1500.
Molteni, F., Buizza, R., Palmer, T.N., Petroliagis, T., 1996. The ECMWF ensemble prediction system: methodology and validation. Quarterly Journal of the Royal Meteorological Society 122 (529), 73–119.
Montmerle, T., Rabier, F., Fischer, C., 2007. Relative impact of polar-orbiting and geostationary satellite radiances in the Aladin/France numerical weather prediction system. Quarterly Journal of the Royal Meteorological Society 133 (624), 655–671.
Morss, R.E., Wilhelmi, O.V., Downton, M.W., Gruntfest, E., 2005. Flood risk, uncertainty, and scientific information for decision making — lessons from an interdisciplinary project. Bulletin of the American Meteorological Society 86 (11), 1593–1601.
Morss, R.E., Demuth, J.L., Lazo, J.K., 2008. Communicating uncertainty in weather forecasts: a survey of the US public. Weather and Forecasting 23 (5), 974–991.
Mueller, M., Tinz, M., Assmann, A., Krahe, P., Rachimow, C., Daamen, K., Bliefernicht, J., Ebert, C., Kunz, M., Schipper, J.W., Meinel, G., Hennersdorf, J., 2009. Short-range plain flood forecasting and risk management in the Bavarian Danube basin. In: Samuels, P., Huntington, S., Allsop, W., Harrop, J. (Eds.), Flood Risk Management: Research and Practice, pp. 1127–1134.
Ntelekos, A.A., Georgakakos, K.P., Krajewski, W.F., 2006. On the uncertainties of flash flood guidance: toward probabilistic forecasting of flash floods. Journal of Hydrometeorology 7 (5), 896–915.
Orlandi, A., Ortolani, A., Meneguzzo, F., Levizzani, V., Torricella, F., Turk, F.J., 2004. Rainfall assimilation in RAMS by means of the Kuo parameterisation inversion: method and preliminary results. Journal of Hydrology 288 (1–2), 20–35.
Palmer, T.N., 2001. A nonlinear dynamical perspective on model error: a proposal for non-local stochastic–dynamic parametrization in weather and climate prediction models. Quarterly Journal of the Royal Meteorological Society 127 (572), 279–304.
Palmer, T.N., Buizza, R., 2007. Fifteenth anniversary of EPS. ECMWF Newsletter 114, 14.
Palmer, T.N., Doblas-Reyes, F.J., Hagedorn, R., Weisheimer, A., 2005. Probabilistic prediction of climate using multi-model ensembles: from basics to applications. Philosophical Transactions of the Royal Society B: Biological Sciences 360 (1463), 1991–1998.
Pappenberger, F., Beven, K.J., 2004. Functional classification and evaluation of hydrographs based on multicomponent mapping. International Journal of River Basin Management 2 (2), 89–100.
Pappenberger, F., Beven, K.J., Hunter, N.M., Bates, P.D., Gouweleeuw, B.T., Thielen, J., de Roo, A.P.J., 2005. Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall–runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS). Hydrology and Earth System Sciences 9 (4), 381–393.
Pappenberger, F., Harvey, H., Beven, K.J., Hall, J., Romanowicz, R., Smith, P.J., 2006. Risk and Uncertainty — Tools and Implementation. UK Flood Risk Management Research Consortium, Lancaster University, Lancaster.
Pappenberger, F., Frodsham, K., Beven, K., Romanowicz, R., Matgen, P., 2007. Fuzzy set approach to calibrating distributed flood inundation models using remote sensing observations. Hydrology and Earth System Sciences 11 (2), 739–752.
Pappenberger, F., Scipal, K., Buizza, R., 2008. Hydrological aspects of meteorological verification. Atmospheric Science Letters 9 (2), 43–52.
Parent du Châtelet, J., Tabary, P., Gueguen, C., Fradon, B., 2006. The Meteo-France Single-radar and Composite QPE Operational Products. ERAD Publication Series, pp. 299–302.
Park, Y.-Y., Buizza, R., Leutbecher, M., 2008. TIGGE: preliminary results on comparing and combining ensembles. Quarterly Journal of the Royal Meteorological Society 134 (637), 2029–2050.
Parker, D., Fordham, M., 1996. An evaluation of flood forecasting, warning and response systems in the European Union. Water Resources Management 10 (4), 279–302.
Pegram, G., Llort, X., Sempere-Torres, D., 2011. Radar-rainfall: separating signal and noise fields to generate meaningful ensembles. Atmospheric Research 100, 226–236 (this issue).
Ranzi, R., Bacchi, B., Grossi, G., 2003. Runoff measurements and hydrological modelling for the estimation of rainfall volumes in an Alpine basin. Quarterly Journal of the Royal Meteorological Society 129 (588), 653–672.
Ranzi, R., Zappa, M., Bacchi, B., 2007. Hydrological aspects of the Mesoscale Alpine Programme: findings from field experiments and simulations. Quarterly Journal of the Royal Meteorological Society 133 (625), 867–880.
Reitebuch, O., Volkert, H., Werner, C., Dabas, A., Delville, P., Drobinski, P., Flamant, P.H., Richard, E., 2003. Determination of airflow across the

Alpine ridge by a combination of airborne Doppler lidar, routine radiosounding and numerical simulation. Quarterly Journal of the Royal Meteorological Society 129 (588), 715–727.
Richard, E., Buzzi, A., Zangl, G., 2007. Quantitative precipitation forecasting in the Alps: the advances achieved by the Mesoscale Alpine Programme. Quarterly Journal of the Royal Meteorological Society 133 (625), 831–846.
Romang, H., Zappa, M., Hilker, N., Gerber, M., Dufour, F., Frede, V., Bérod, D., Oplatka, M., Rhyner, J., 2011. IFKIS-Hydro: an early warning and information system for floods and debris flows. Natural Hazards 56 (2), 509–527 (19).
Rossa, A., Bruen, M., Fruehwald, D., Macpherson, B., Holleman, I., Michelson, D., Michaelides, S., 2005a. COST 717 Action — Use of Radar Observation in Hydrology and NWP Models. COST Meteorology EUR 21954. COST, 286 pp.
Rossa, A., Bruen, M., Collier, C., Frühwald, D., Germann, U., Gjertsen, U., Macpherson, B., Michelson, D., Rotach, M.W. and Sempere-Torres, D., 2005. COST 731, Propagation of uncertainty in advanced meteo-hydrological forecast systems, Memorandum of Understanding, available at http://www.cost.eu.
Rossa, A., Haase, G., Keil, C., Alberoni, P., Ballard, S., Bech, J., Germann, U., Pfeifer, M., Salonen, K., 2010a. Propagation of uncertainty from observing systems into NWP: COST-731 Working Group 1. Atmospheric Science Letters 11 (2), 145–152.
Rossa, A., Laudanna Del Guerra, F., Borga, M., Zanon, F., Settin, T., Leuenberger, D., 2010b. Radar-driven high-resolution hydro-meteorological forecasts of the 26 September 2007 Venice Flash Flood. Journal of Hydrology 394, 230–244. doi:10.1016/j.jhydrol.2010.08.035.
Rotach, M.W., Ambrosetti, P., Ament, F., Appenzeller, C., Arpagaus, M., Bauer, H.-S., Behrendt, A., Bouttier, F., ois, Buzzi, A., Corazza, M., Davolio, S., Denhard, M., Dorninger, M., Fontannaz, L., Frick, J., Fundel, F., Germann, U., Gorgas, T., Hegg, C., Hering, A., Keil, C., Liniger, M.A., Marsigli, C., McTaggart-Cowan, R., Montaini, A., Mylne, K., Ranzi, R., Richard, E., Rossa, A., Santos, M., oz, D., Sch, r, C., Seity, Y., Staudinger, M., Stoll, M., Volkert, H., Walser, A., Wang, Y., Werhahn, J., Wulfmeyer, V., Zappa, M., 2009. MAP D-PHASE: real-time demonstration of weather forecast quality in the Alpine Region. Bulletin of the American Meteorological Society 90 (9), 1321–1336.
Roulin, E., 2007. Skill and relative economic value of medium-range hydrological ensemble predictions. Hydrology and Earth System Sciences 11 (2), 725–737.
Saltikoff, E., Gjertsen, U., Michelson, D., Holleman, I., Seltmann, J., Odakivi, K., Huuskonen, A., Hohti, H., Koistinen, J., Pohjola, H., Haase, G., 2004. Radar Data Quality Issues in Northern Europe. ERAD Publication Series, pp. 212–215.
Samuels, P., Klijn, F., Dijkman, J., 2006. An analysis of the current practice of policies on river flood risk management in different countries. Irrigation and Drainage 55, S141–S150.
Schaake, J., Larson, L., Ams, A.M.S., 1998. Ensemble streamflow prediction (ESP): progress and research needs. Special Symposium on Hydrology J19–J24.
Schaake, J., Demargne, J., Hartman, R., Mullusky, M., Welles, E., Wu, L., Herr, H., Fan, X., Seo, D.J., 2007a. Precipitation and temperature ensemble forecasts from single-value forecasts. Hydrology and Earth System Sciences Discussions 4 (2), 655–717.
Schaake, J., Hamill, T.M., Buizza, R., Clark, M., 2007b. HEPEX: The Hydrological Ensemble Prediction Experiment. Bulletin of the American Meteorological Society 88 (10), 1541–1547.
Schröter, K., Llort, X., Velasco-Forero, C., Ostrowski, M., Sempere-Torres, D., 2011. Implications of radar rainfall estimates uncertainty on distributed hydrological model predictions. Atmospheric Research. 100, 237–245 (this issue).
Seed, A.W., 2003. A dynamic and spatial scaling approach to advection forecasting. Journal of Applied Meteorology 42 (3), 381–388.
Sevruk, B., 1996. Adjustment of tipping-bucket precipitation gauge measurements. Atmospheric Research 42 (1–4), 237–246.
Stephan, K., Klink, S., Schraff, C., 2008. Assimilation of radar-derived rain rates into the convective-scale model COSMO-DE at DWD. Quarterly Journal of the Royal Meteorological Society 134 (634), 1315–1326.
Sun, J., 2005. Convective-scale assimilation of radar data: progress and challenges. Quarterly Journal of the Royal Meteorological Society 131 (613), 3439–3463.
Sun, J., Crook, N.A., 1998. Dynamical and microphysical retrieval from Doppler radar observations using a cloud model and its adjoint. Part II: retrieval experiments of an observed Florida convective storm. Journal of the Atmospheric Sciences 55 (5), 835–852.
Szturc, J., Osrodka, K., Jurczyk, A., Jelonek, L., 2008. Concept of dealing with uncertainty in radar-based data for hydrological purpose. Natural Hazards and Earth System Sciences 8 (2), 267–279.
Thielen, J., Schaake, J., Hartman, R., Buizza, R., 2008. Aims, challenges and progress of the Hydrological Ensemble Prediction Experiment (HEPEX) following the third HEPEX workshop held in Stresa 27 to 29 June 2007. Atmospheric Science Letters 9 (2), 29–35.
Thielen, J., Bartholmes, J., Ramos, M.-H., de Roo, A., 2009. The European Flood Alert System — part 1: concept and development. Hydrology and Earth System Sciences 13 (2), 125–140.
Todini, E., 2004. Role and treatment of uncertainty in real-time flood forecasting. Hydrological Processes 18 (14), 2743–2746.
Todini, E., Mantovan, P., 2007. Comment on: 'On undermining the science?' by Keith Beven. Hydrological Processes 21 (12), 1633–1638.
Uijlenhoet, R., Berne, A., 2008. Stochastic simulation experiment to assess radar rainfall retrieval uncertainties associated with attenuation and its correction. Hydrology and Earth System Sciences 12 (2), 587–601.
van der Keur, P., Iversen, B.V., 2006. Uncertainty in soil physical data at river basin scale — a review. Hydrology and Earth System Sciences 10 (6), 889–902.
Vehvilainen, B., Huttunen, M., Huttunen, I., 2005. Hydrological forecasting and real time monitoring in Finland: the watershed simulation and forecasting system (WSFS), Innovation. Advances and Implementation of Flood Forecasting Technology, Tromso, Norway. 17–19 October 2005.
Viviroli, D., Zappa, M., Gurtz, J., Weingartner, R., 2009. An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools. Environmental Modelling & Software 24 (10), 1209–1222.
Volkert, H., Gutermann, T., 2007. Inter-domain cooperation for mesoscale atmospheric laboratories: the Mesoscale Alpine Programme as a rich study case. Quarterly Journal of the Royal Meteorological Society 133 (625), 949–967.
Vrugt, J.A., Diks, C.G.H., Gupta, H.V., Bouten, W., Verstraten, J.M., 2005. Improved treatment of uncertainty in hydrologic modeling: combining the strengths of global optimization and data assimilation. Water Resources Research 41 (1).
Wagener, T., McIntyre, N., Lees, M.J., Wheater, H.S., Gupta, H.V., 2003. Towards reduced uncertainty in conceptual rainfall–runoff modelling: dynamic identifiability analysis. Hydrological Processes 17 (2), 455–476.
Walser, A., Schär, C., 2004. Convection-resolving precipitation forecasting and its predictability in Alpine river catchments. Journal of Hydrology 288 (1–2), 57–73.
Walser, A., Lüthi, D., Schär, C., 2004. Predictability of precipitation in a cloud-resolving model. Monthly Weather Review 132 (2), 560–577.
Werner, M., Cranston, M., 2009. Understanding the value of radar rainfall nowcasts in flood forecasting and warning in flashy catchments. Meteorological Applications 16 (1), 41–55.
Werner, M., Cranston, M., Harrison, T., Whitfield, D., Schellekens, J., 2009. Recent developments in operational flood forecasting in England, Wales and Scotland. Meteorological Applications 16 (1), 13–22.
Wetterhall, F., Halldin, S., Xu, C.Y., 2005. Statistical precipitation downscaling in central Sweden with the analogue method. Journal of Hydrology 306 (1–4), 174–190.
Weygandt, S., Loughe, A.F., Benjamin, S.G., Mahoney, J.L., 2004. Scale Sensitivities in Model Precipitation Skill Scores during IHOP. In: American Meteorological Society (Ed.), 22nd Conference on Severe Local Storms, Hyannis, MA.
Wilks, D.S., 2006. Statistical Methods in the Atmospheric Sciences. Elsevier, Amsterdam. 627 pp.
Xuan, Y., Cluckie, I.D., Wang, Y., 2009. Uncertainty analysis of hydrological ensemble forecasts in a distributed model utilising short-range rainfall prediction. Hydrology and Earth System Sciences 13 (3), 293–303.
Zappa, M., Rotach, M.W., Arpagaus, M., Dorninger, M., Hegg, C., Montani, A., Ranzi, R., Ament, F., Germann, U., Grossi, G., Jaun, S., Rossa, A., Vogt, S., Walser, A., Wehrhan, J., Wunram, C., 2008. MAP D-PHASE: real-time demonstration of hydrological ensemble prediction systems. Atmospheric Science Letters 9 (2), 80–87.
Zappa, M., Beven, K.J., Bruen, M., Cofino, A., Kok, K., Martin, E., Nurmi, P., Orfila, B., Roulin, E., Schröter, K., Seed, A., Stzurc, J., Vehviläinen, B., Germann, U., Rossa, A., 2010. Propagation of uncertainty from observing systems and NWP into hydrological models: COST-731 Working Group 2. Atmospheric Science Letters 11 (2), 83–91.
Zappa, M., Jaun, S., Germann, U., Walser, A., Fundel, F., 2011. Superposition of three sources of uncertainties in operational flood forecasting chains. Atmospheric Research 100, 246–262 (this issue).

II

Flood nowcasting in the southern Swiss Alps using radar ensemble

KATHARINA LIECHTI[1], FELIX FUNDEL[1], URS GERMANN[2] & MASSIMILIANO ZAPPA[1]

1 *Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland*
kaethi.liechti@wsl.ch

2 *Federal Office for Meteorology and Climatology, MeteoSwiss, Via Monti 146, CH-6605 Locarno, Switzerland*

Abstract Since April 2007 the MeteoSwiss radar ensemble product REAL has been in operation and used for operational flash flood nowcasting by the WSL. REAL is computed for an area in the southern Swiss Alps where orographic and convective precipitation is frequent. These ensemble QPEs are processed by the semi-distributed hydrological model PREVAH. This provides operational ensemble nowcasts for several basins with areas from 44 to 1500 km^2 prone to flash floods and floods, respectively. Performances of discharge nowcasts driven by REAL are compared to performances of nowcasts forced by deterministic radar QPE and by interpolated raingauge data. We show that REAL outperforms deterministic radar QPE over the whole range of discharges, while the intercomparison with interpolated raingauge data is threshold dependent. Further we show that even though REAL nowcasts are underdispersive they have skill and can be a valuable means to produce hydrological nowcasts especially in ungauged catchments.

Key words radar ensemble; nowcasting; flash flood; flood; probabilistic; verification

INTRODUCTION

Mountainous catchments are often prone to flash floods, as their topography favours heavy convective precipitation events (Panziera & Germann, 2010). Due to shallow soils and steep slopes the runoff response time is generally short. Observations covering remote regions are typically scarce. Quantitative radar precipitation estimates (QPEs) are sometimes the only available information about precipitation in a remote area and provide valuable additional information about the spatial distribution of precipitation. However, uncertainties in radar QPEs for Alpine regions are large because of severe shielding of the radar beam by mountains, orographic precipitation mechanisms not fully seen by the radar, and strong mountain returns (clutter) (Germann *et al*., 2006b). A promising solution to express the residual uncertainty in radar QPEs is to generate an ensemble of precipitation fields (e.g. Krajewski & Georgakakos, 1985). An example can be found in Aghakouchak *et al*. (2010), where three different remotely-sensed rainfall ensemble generators are compared.

The radar ensemble generator used in this study was developed at MeteoSwiss and was coupled to a hydrological runoff model (Germann *et al*., 2009; Zappa *et al*., 2011). This resulted in the first operationally running hydro-meteorological model chain using a radar rainfall ensemble and providing hydrological nowcasts at hourly time steps in a mountainous region (Zappa *et al*., 2008).

During the 4 years of operational experience with this novel model chain, a unique dataset consisting of various types of input data (e.g. radar ensemble, deterministic radar, interpolated raingauge data, numerical weather predictions) and their resulting hydrological simulations could be acquired for validation.

In this study we present a probabilistic assessment of the performance of the radar ensemble (REAL) driven discharge nowcasts in comparison to nowcasts driven by interpolated raingauge data and deterministic radar QPEs. The analysed period ranges from April 2007 to December 2009.

STUDY AREA AND DATA

Test site

Alpine catchments are ideal sites to test a radar ensemble in combination with a hydrological runoff model. Due to their topography, persistent orographic precipitation combined with rapid runoff generation often leads to flash flood events associated with considerable damage (Germann *et al*., 2009). Our experiments are focused on an area in the southern Swiss Alps including the four test catchments Pincascia, Verzasca, Maggia and Ticino, considered in this study (Fig. 1). The Pincascia with its 44 km^2 is a sub-catchment of the Verzasca. The Verzasca with an area of 186 km^2 down to the gauge in Lavertezzo is little affected by human activities. The Maggia catchment down to Locarno (926 km^2) is influenced by reservoir lakes in the upper part of the catchment for hydropower production. The Ticino catchment down to Bellinzona encompasses an area of 1515 km^2 and is also affected by water management in connection with hydropower production. Major infrastructure (highway, railway) is situated in the main valley.

All the catchments are characterised by snowmelt in spring and early summer and heavy rainfall events in autumn. Elevations for Pincascia, Verzasca, Maggia and Ticino range from 540 to 2500, 490 to 2900, 200 to 3300 and 220 to 3400 m a.m.s.l., respectively.

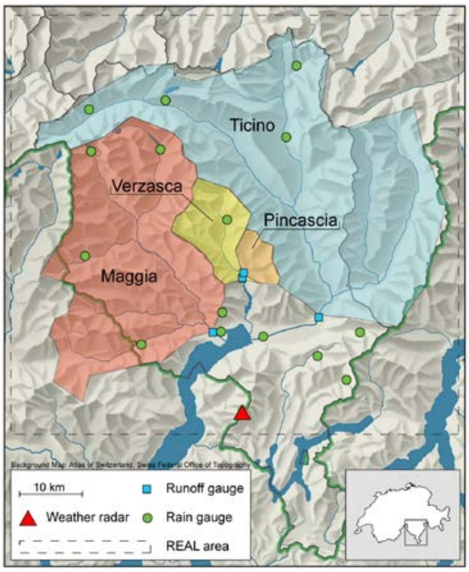

Fig. 1 REAL area with test catchments, raingauges and weather radar location.

REAL and other data

Weather radar offers the possibility to estimate precipitation at a high space and time resolution. But there are many sources of error that have to be dealt with. MeteoSwiss developed and implemented a series of algorithms to correct for several of the many errors inherent in radar reflectivity measurements. Although significant improvements were achieved, the residual uncertainty for hydrological applications is still relatively large (Germann *et al*., 2006a).

The generation of an ensemble of radar precipitation fields is an elegant way to express the residual uncertainty in radar QPEs (Germann *et al*., 2009). For the MeteoSwiss radar ensemble (REAL) the radar precipitation field is perturbed with correlated random noise. In this way the residual space–time uncertainty in the radar estimates is taken into account. The perturbation fields are generated by combining stochastic simulation techniques with detailed knowledge on the space-time variance and auto-covariance of radar errors (Germann *et al*., 2006a). The radar ensemble is generated hourly with a spatial resolution of 2 km. The number of ensemble members is set to 25 (Germann *et al*., 2009).

The meteorological data to run the hydrological model (air temperature, water vapour pressure, global radiation, wind speed, sunshine duration and precipitation) and the deterministic radar QPEs (1-km^2 spatial resolution) are also provided by MeteoSwiss. The raingauge data are interpolated over the test areas with inverse-distance weighting. The discharge measurements for verification are available in hourly time steps up to December 2009. For analysis, all the resulting nowcasts and the observed discharge time series were aggregated to series of daily maxima. This results in a database of 997 days for evaluation.

METHOD

Model chain

REAL is coupled to the semi-distributed hydrological model PREVAH (Viviroli *et al*., 2009). Calibration and verification of PREVAH are presented in Wöhling *et al*. (2006) and Ranzi *et al*. (2007). This model chain produces operational nowcasts at hourly time steps ever since April 2007. For the period with available validated discharge data the model chain was re-run using REAL, interpolated raingauge data and deterministic radar QPEs as precipitation input (Fig. 2).

It is our goal to analyse the development of spread of the REAL ensemble QPEs over time. In numerical weather prediction, spread increases with lead time (Bartholmes *et al*., 2009). By analogy, the spread of radar ensemble products is expected to change with the number of hours the precipitation ensemble is used for the generation of initial states for a subsequent nowcast.

We allow each of the 25 REAL members to build up a 10-day chain of spatially and temporally correlated precipitation values. Thus spread can develop from day to day. During long dry periods the spread can converge. During long wet periods the spread can grow. Our set-up starts from "day minus 10" with identical initial conditions for the hydrological simulations. In case of rainfall the 25 members of weather radar precipitation propagate separately through the hydrological model.

We repeated the 10-day simulations starting them consecutively each day from April 2007 until December 2009. This set-up allows us to create chains of discharge values where the forcing REAL-precipitation input has identical spread development time. These chains are evaluated with standard deterministic and probabilistic verification metrics as generally used for analysing ensemble discharge forecasts for different lead times (e.g. Jaun & Ahrens, 2009; Addor *et al*., 2011).

Performance measures

For comparison of the performance of the deterministic nowcasts driven by interpolated raingauge data and radar QPEs and the probabilistic nowcasts driven by REAL, the Brier Skill Score (BSS) was chosen. The BSS is the most common measure for the verification of probabilistic forecasts of dichotomous events and allows a direct comparison of deterministic and ensemble predictions (Wilks, 2006). The BSS is based on the Brier Score (BS), which is essentially the mean squared error of the probability forecasts, given a dichotomous event (exceedence of a threshold or not), defined as:

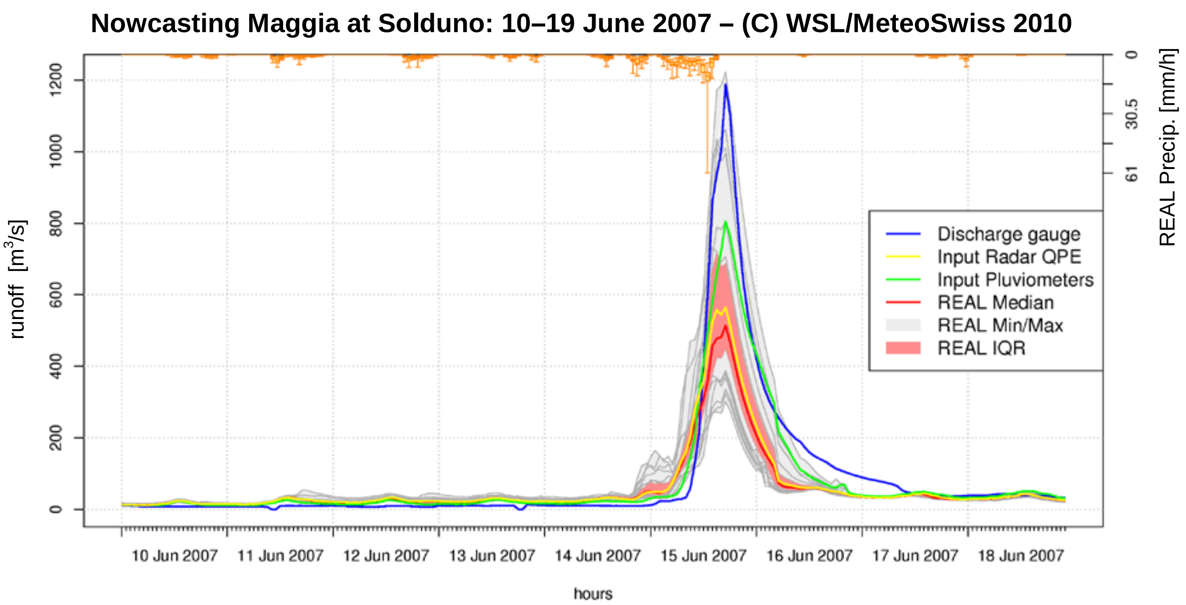

Fig. 2 Example of a REAL nowcast and *a posteriori* observed discharge (blue).

$$\mathrm{BS} = \frac{1}{n}\sum_{d=1}^{n}(y_d - o_d)^2 \tag{1}$$

where y_d is the forecast probability of threshold exceedence (between 0 and 1) and o_d is the observed outcome (1 if threshold is exceeded, 0 if not) for forecast *d* of *n* assessed. The BSS is the improvement of the forecast in BS over that for a reference forecast BS_{Cl} (Wilks, 2006) such that:

$$\mathrm{BSS} = 1 - \frac{\mathrm{BS}}{\mathrm{BS}_{Cl}}. \tag{2}$$

In this study the reference forecast is the climatological probability of exceedence of the predefined threshold. A perfect forecast has a skill of 1, whereas forecasts worse than the climatological forecast have a skill below 0.

The false alarm ratio (FAR) and the probability of detection (POD) are measures for deterministic predictions, and therefore the REAL ensemble was reduced to its median. The FAR is the fraction of positive forecasts (exceedence of threshold) that turn out to be wrong. The best FAR value is zero and the worst is one.

The POD is the ratio of correctly forecasted threshold exceedences to the number of times it actually happened. The best POD value is one and the worst is zero. The POD can always be improved by forecasting the event more often; however this usually results in higher FAR and for extreme events results in an overforecasting bias (Bartholmes *et al*., 2009). So both the POD and FAR are shown.

Additionally, rank histograms (RH) provide information about the spread and bias of the REAL ensemble. All performance measures are calculated on the 0.8 and 0.95 quantile of the total available discharge time series for the different catchments (Ticino and Maggia 36 years, Verzasca 21 years and Pincascia 18 years).

RESULTS

The deterministic nowcasts driven by interpolated raingauge data and deterministic radar data are the same for all chains, which implies that the BSS, FAR and POD values are constant over the 10 chains (Fig. 3). The performance of the probabilistic REAL nowcasts vary over the 10 chains. The difference between the chains is however relatively small (Fig. 3).

BSS All nowcasts (except Maggia 0.8 quantile) show positive BSS values. The BSS values for the 0.95 quantile are higher than for the 0.8 quantile for all catchments and data types. From chain 8 to chain 10 BSS values for REAL decrease for all catchments. REAL outperforms deterministic radar over all thresholds, whereas a comparison to the performance of raingauge data is threshold dependent.

FAR For the 0.8 quantile the FAR achieved by raingauge data, deterministic radar and REAL lie very close together in all catchments. Very low FAR values are obtained for the Pincascia catchment, whereas the FAR for the Maggia is comparatively high (0.3–0.4). For Verzasca and Ticino FAR values are slightly higher for the 0.95 quantile than for the 0.8 quantile. For all catchments the REAL FAR increases from chain 8 to 10.

POD The POD for Verzasca, Maggia and Ticino are very similar. For the two threshold quantiles they lie between 0.6–0.7 and 0.7–0.9, respectively. From chain 8 to 10 REAL POD decreases in all catchments. Values for Pincascia are low compared to the other catchments.

Rank histogram The rank histograms show an underdispersion of the REAL ensemble for all catchments, most clearly for the 0.8 quantile. Additionally they show a clear underprediction for the Pincascia for both thresholds. The REAL ensembles for Verzasca and Ticino have a good average spread for the 0.95 quantile (Fig. 4).

DISCUSSION

The positive BSS values show that there is additional skill over a climatological forecast for all data types. The superiority of the REAL performance over the performance of deterministic radar

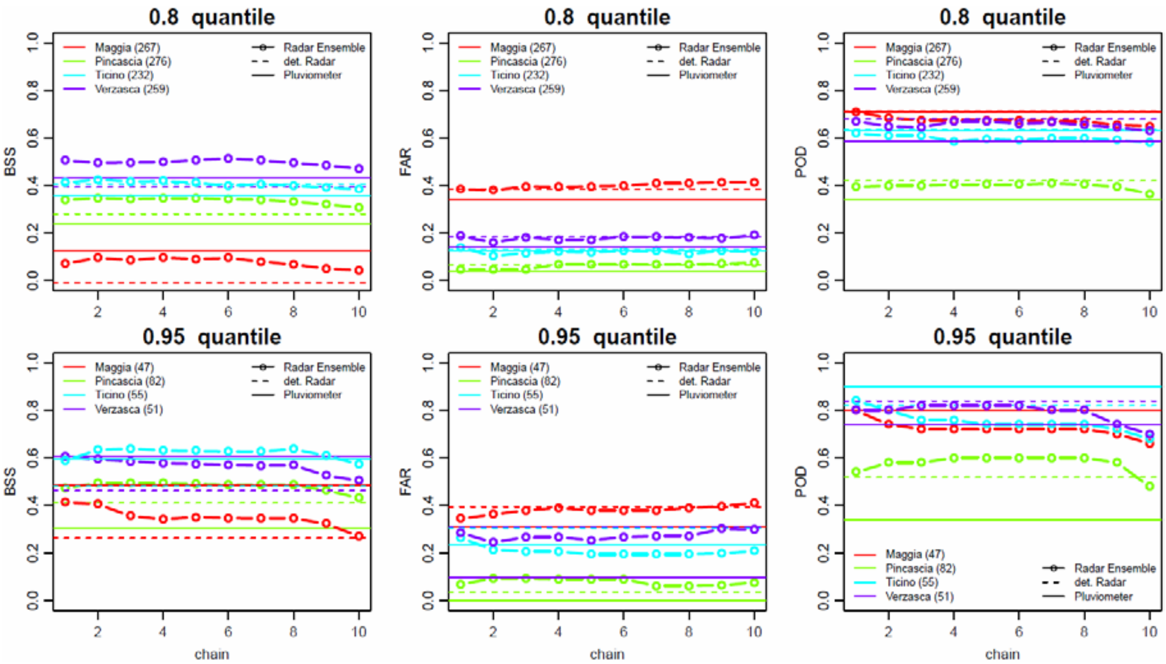

Fig. 3 Brier Skill Score (BSS), False alarm ratio (FAR) and Probability of Detection (POD) for the four test catchments calculated for the different precipitation input and the 0.8 and 0.95 discharge quantiles. The numbers in the brackets are the number of events reaching or exceeding the respective quantile.

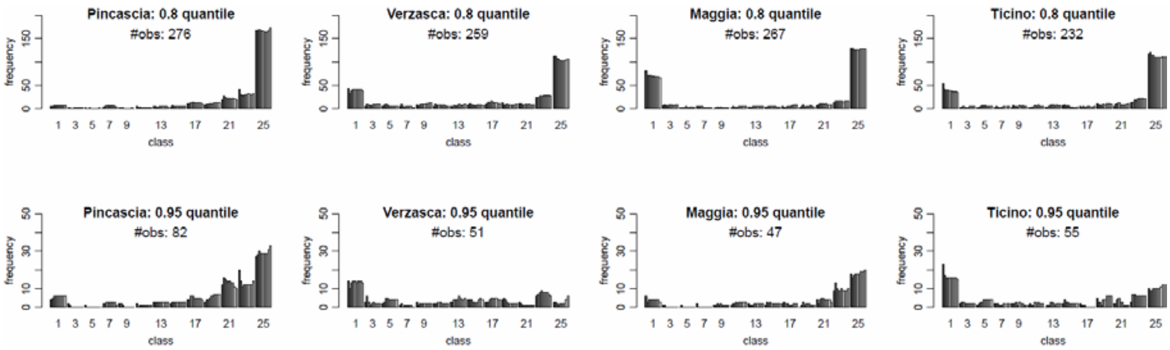

Fig. 4 Rank histograms of the REAL ensemble for the 0.8 and 0.95 quantiles. The thin columns inside each class represent different chains.

QPEs in all catchments is a clear argument for the probabilistic approach. An optimal spread development time for REAL cannot be deduced from the BSS, as the difference between the chains is small. However, the BSS, FAR and POD values all show a decline of performance from chain 8 to 10. This indicates that a spread development time of more than 8 days is not useful.

The comparison of the BSS values for the Verzasca and Pincascia catchments show the value of REAL for ungauged catchments. While in the case of the Verzasca for the 0.95 quantile the raingauge data achieves higher skill than REAL, it is clearly the opposite for the ungauged sub-catchment Pincascia (Figs 1 and 3). For the Pincascia the low FAR and the relatively low POD values are remarkable. This indicates that forecasted events exceeding the thresholds shown in Fig. 3 are reliable and thus will most probably occur. On the other hand, a big part of the events is missed by the forecasts. Due to the lack of a raingauge in the catchment, this is most pronounced for the POD values achieved by interpolated raingauge data. The same signal can be seen in the rank histograms which clearly show an underpredicting bias for the Pincascia for both thresholds.

The results for the Maggia catchment mirror the influence of water management for hydropower production. Precipitation of heavy storm events is often stored in the retention lakes and does not contribute to discharge. This results in low BSS and FAR and high POD values.

It is rather surprising that the scores for the 0.95 quantile are better than for the 0.8 quantile. A seasonal verification or a seasonal varying threshold may change this result. Finally it has to be acknowledged that the presented statistics rely on a series of 33 months only, which strongly limits the number of events exceeding the quantiles relevant for flash flood and flood events in these mountainous catchments.

CONCLUSIONS AND OUTLOOK

For catchments with sparse raingauge networks, REAL is a valuable solution to produce discharge nowcasts. We showed that REAL nowcasts outperform discharge nowcasts forced by deterministic radar QPEs on all thresholds. The REAL ensemble is underdispersive but shows skill nonetheless. We showed that the performance of nowcasts forced by REAL decreases after a spread development time of 8 days.

REAL can be used to generate initial conditions for a subsequent initialisation of the hydrological model with, for example, data from numerical weather prediction models (NWP; Zappa *et al*., 2011). Such model coupling has already been tested with probabilistic and deterministic NWP products (COSMO-LEPS, COSMO-7, COSMO-2) for forecasts with lead times of a few days. Panziera & Germann (2010) describe a new radar ensemble product tailored for radar QPEs with short lead-times. Such products have the potential to enhance the reliability of nowcasts and short-term forecasts in mountainous areas and will soon be tested for hydrological applications.

Acknowledgements The experiments for the Pincascia and Verzasca and three of the authors are funded by the EU FP7 Project IMPRINTS (grant agreement no. 226555 / FP7-ENV-2008-1-226555). Experiments for the Maggia and Ticino basins are supported by a grant from the administration of Cantone Ticino through the Interreg IV Project FLORA. Discharge data are provided by the Swiss Federal office for the environment.

REFERENCES

Addor, N., Jaun, S. & Zappa, M. (2011) An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): skill, case studies and scenarios. *Hydrol. Earth Syst. Sci*. 15, 2327–2347.

Aghakouchak, A., Habib, E. & Bardossy, A. (2010) A comparison of three remotely sensed rainfall ensemble generators. *Atmos. Res*. 98(2-4), 387–399.

Bartholmes, J. C., Thielen, J., Ramos, M. H. & Gentilini, S. (2009) The European Flood Alert System EFAS – Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts. *Hydrol. Earth Syst. Sci*. 13(2), 141–153.

Germann, U., Berenguer, M., Sempere-Torres, D. & Salvadè, G. (2006) Ensemble radar precipitation estimation — a new topic on the radar horizon. In: *4th European Conference on Radar in Meteorology and Hydrology* (Barcelona), 559–562.

Germann, U., Berenguer, M., Sempere-Torres, D. & Zappa, M. (2009) REAL – Ensemble radar precipitation estimation for hydrology in a mountainous region. *Quart. J. Roy. Met. Soc.* 135(639), 445–456.

Germann, U., Galli, G., Boscacci, M. & Bolliger, M., 2006. Radar precipitation measurement in a mountainous region. *Quart. J. Roy. Met. Soc.* 132(618), 1669–1692.

Jaun, S. & Ahrens, B. (2009) Evaluation of a probabilistic hydrometeorological forecast system. *Hydrol. Earth Syst. Sci. Discuss.* 6(2), 1843–1877.

Krajewski, W. F. & Georgakakos, K. P. (1985) Synthesis of radar rainfall data. *Water Resour. Res.* 21(5), 764–768.

Panziera, L. & Germann, U. (2010) The relation between airflow and orographic precipitation on the southern side of the Alps as revealed by weather radar. *Q. J. Roy. Meteor. Soc.* 136(646), 222–238.

Ranzi, R., Zappa, M. & Bacchi, B. (2007) Hydrological aspects of the Mesoscale Alpine Programme: Findings from field experiments and simulations. *Q. J. Roy. Meteor. Soc.* 133(625), 867–880.

Viviroli, D., Zappa, M., Gurtz, J. & Weingartner, R. (2009) An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools. *Environ. Modell. Softw.* 24(10), 1209–1222.

Wilks, D. S. (2006) *Statistical Methods in the Atmospheric Sciences*. Elsevier, Amsterdam, 627 pp.

Wöhling, T., Lennartz, F. & Zappa, M. (2006) Technical Note: updating procedure for flood forecasting with conceptual HBV-type models. *Hydrol. Earth Syst. Sci.* 10(6), 783–788.

Zappa, M., Jaun, S., Germann, U., Walser, A. & Fundel, F. (2011) Superposition of three sources of uncertainties in operational flood forecasting chains. *Atmos. Res.* 100(2–3), 246–262.

Zappa, M., Rotach, M. W., Arpagaus, M., Dorninger, M., Hegg, C., Montani, A., Ranzi, R., Ament, F., Germann, U., Grossi, G., Jaun, S., Rossa, A., Vogt, S., Walser, A., Wehrhan, J. & Wunram, C. (2008) MAP D-PHASE: real-time demonstration of hydrological ensemble prediction systems. *Atmos. Sci. Lett.* 9(2), 80–87.

III

HYDROLOGICAL PROCESSES
Hydrol. Process. (2012)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/hyp.9458

Probabilistic evaluation of ensemble discharge nowcasts in two nested Alpine basins prone to flash floods

Katharina Liechti,[1]* Massimiliano Zappa,[1] Felix Fundel[1] and Urs Germann[2]

[1] *Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland*
[2] *Federal Office for Meteorology and Climatology, MeteoSwiss, Locarno, Switzerland*

Abstract:

The quality of hydrological discharge simulations depends to a great extent on the uncertainties in the meteorological input and the model parameterization. To quantify these uncertainties, we adopt ensemble techniques in a 4-year nowcast experiment for two nested flash-flood–prone basins in the southern Swiss Alps.
The spatiotemporal uncertainties in the weather radar quantitative precipitation estimates (QPE) were accounted for by applying an ensemble of 25 radar fields. To account for uncertainties in model parameterization, a Monte Carlo experiment was run to find 26 equifinal model realizations. The resulting parameter ensemble, consisting of 26 members, was run with precipitation input obtained from interpolated pluviometer data and with the deterministic operational weather radar QPE.
To produce the discharge nowcast, the PREcipitation-Runoff-EVApotranspiration HRU Model (PREVAH) was used. PREVAH was calibrated for the main catchment Verzasca. The results for the sub-catchment Pincascia are an independent internal verification of the nowcasting system.
The three ensemble nowcasts and the two deterministic nowcasts are evaluated for a 4-year time series and for two events included in that period. The event analysis shows no clear superiority for either pluviometer-based or radar-based nowcasts. The performance for single events depends heavily on the storm characteristics. However, the evaluation of the 4-year nowcast shows that pluviometer-based nowcasts outperform radar-based nowcasts in the gauged and calibrated catchment and that there is added value in the application of parameter ensembles.
For the small, ungauged catchment, the results achieved by the radar-based nowcasts are superior to the pluviometer-based nowcasts. Especially the radar ensemble proves to be of significant advantage for flash flood nowcasts in such catchments.

KEY WORDS flash flood; nowcast; ensemble; verification; probabilistic

Received 10 November 2011; Accepted 31 May 2012

INTRODUCTION

Flash foods are among the most serious natural hazards, as they frequently cause severe damage to the environment and infrastructure, as well as loss of life (Chiang *et al.*, 2007; Borga *et al.*, 2011; Hapuarachchi *et al.*, 2011). Mountainous catchments are particularly prone to flash floods, as their topography favours heavy convective precipitation events (Smith, 1979; Panziera and Germann, 2010). The shallow soils and steep slopes in such catchments mean that the response time is generally short (Barredo, 2007). Those flashy catchments are usually small in size and often poorly gauged or ungauged, which makes it very difficult to obtain reliable estimates of precipitation (Werner and Cranston, 2009; Hapuarachchi *et al.*, 2011).

These conditions make flash flood forecasting one of the most challenging tasks in operational hydrology (Collier, 2007). In the past, flood forecasting mostly involved using ground-based gauge observations. For remote regions, however, where rain gauge measurements are limited, it is difficult to produce reliable discharge forecasts due to the uncertainty in the forcing input precipitation (Chiang *et al.*, 2007). In such cases, remotely sensed data are particularly useful for hydrological modelling (Hapuarachchi *et al.*, 2011). Due to their good space–time resolution, weather radar quantitative precipitation estimates (QPE) can be of great help in flash flood forecasting even though hydrological applications are highly sensitive to bias and scatter in radar rainfall estimates (Germann *et al.*, 2006b; Collier, 2007; Collier, 2009; Werner and Cranston, 2009).

So far, studies on the reliability of radar rainfall data have concentrated either on the evaluation of selected events (Gourley *et al.*, 2010; Mandapaka *et al.*, 2010; Rossa *et al.*, 2010) or on common verification statistics for long samples of data (Michelson *et al.*, 2005; Krajewski *et al.*, 2010). In most cases, relatively good results have been obtained with radar data, but several problems remain, for instance, the use of radar in mountainous regions (Werner and Cranston, 2009). In the hydrological community, a lack of understanding of the uncertainties in the radar estimates has often led to the exclusion of radar data in hydrological modelling (Rossa *et al.*, 2005). The action COST 731 therefore aimed to address the quantification and communication of the uncertainty in meteorological observation and forecasting,

*Correspondence to: Katharina Liechti, Mountain Hydrology and Mass Movements, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland.
E-mail: kaethi.liechti@wsl.ch

and their effects on hydrological forecasting (Rossa *et al.*, 2011). Several studies have investigated radar uncertainty and its impact on discharge prediction. Germann *et al.* (2009) introduced a radar ensemble to quantify the uncertainty in radar QPE and tested its applicability in mountainous regions with a real-time experiment that coupled the radar ensemble with the semi-distributed rainfall-runoff model PREVAH (Viviroli *et al.*, 2009b). He *et al.* (2011) studied the hydrological impact of the uncertainties in radar QPE by propagating a radar rainfall ensemble through a distributed and integrated water resource model. They analysed 1 year of daily data and found that the resulting hydrological uncertainty was strongly dependent on the scale.

In numerical weather prediction (NWP), the ensemble approach is an established method to quantify uncertainty (Molteni *et al.*, 1996; Zappa *et al.*, 2011). Ensemble prediction systems (EPS) account for the chaotic nature of the atmosphere. In the past few years, this approach has also become established in hydrological forecasting (Cloke and Pappenberger, 2009). Increasing attention has been paid to uncertainty in flash flood forecasts (Villarini *et al.*, 2010), as the deterministic forecasts provide the end users and decision makers with an illusion of certainty (Krzysztofowicz, 2001). The ensemble approach provides a straightforward strategy for application in hydrological flood forecasting as the members of the meteorological ensemble can simply be fed into the hydrological model (Zappa *et al.*, 2010). For a review of ensemble techniques applied in flood forecasting, see Cloke and Pappenberger (2009).

Besides the meteorological input uncertainty, which plays a major role in the outcome of hydrological stream flow simulations (Yatheendradas *et al.*, 2008), the hydrological model uncertainty is a further measure that has to be dealt with (Werner and Cranston, 2009; Zappa *et al.*, 2011). This issue was addressed by Beven and Binley (1992), who introduced the generalized likelihood uncertainty estimation (GLUE) methodology. They suggested that, because there is no perfect model structure and all observations used for model calibration are erroneous to some extent, a true parameter set describing the model cannot be expected, but equally likely parameter sets do exist. The concept of equifinality has since been applied and further developed (Beven and Freer, 2001; Beven, 2006; Choi and Beven, 2007).

Few studies have addressed the combined effects of meteorological input uncertainty and model parameter uncertainty. Zappa *et al.* (2011) investigated the superposition of three sources of uncertainty in operational flood forecasting chains. For seven events, they quantified the combined uncertainty resulting from feeding a hydrometeorological forecasting chain with NWP forecasts (16 members), real-time assimilation of radar precipitation fields (ensemble radar QPE, 25 members), and the equifinal parameter realizations of the hydrological model (parameter ensemble, 26 members). The configuration of the experiment of Zappa *et al.* (2011) was also used in this study.

Unlike in most other studies, we focused on ensemble nowcasting instead of forecasting. Nowcasts are basically very short-term forecasts that typically cover the next 0 to 6 h. This kind of forecast is particularly valuable for flash flood warning systems (Wilson *et al.*, 1998; Berenguer *et al.*, 2005; Panziera *et al.*, 2011). Operational flash flood nowcasting uses radar rainfall estimates and real-time precipitation data from automatic rain gauge stations. These data are sometimes not available, as flashy catchments are often remote and poorly measured or ungauged.

In our study, we evaluated 4 years of continuous discharge nowcasts (2007–2010), conducted in hourly time steps with different ensemble approaches as well as with common deterministic input data for two nested basins: a main catchment including one pluviometer and a sub-catchment including no pluviometer. This allows us to evaluate the nowcast quality produced by different input data over a long time period as well as to internally verify the nowcasting system.

The aim of this study was to investigate if flash flood nowcasts can be improved by applying an ensemble technique to quantify uncertainties in both radar QPE and model parameterization. We also wanted to investigate if either pluviometer-based nowcasts or radar-based nowcasts have an advantage over the other. The internal verification of the data set serves as an assessment of the potential of this methodology for ungauged and poorly gauged catchments.

LOCATION AND DATA

The two catchments chosen for this analysis are located in the southern Swiss Alps (Figure 1). Their steep topography in combination with frequent orographic precipitation in this region makes them prone to flash floods (Panziera *et al.*, 2011). The Pincascia catchment, 44 km^2 in area, is a sub-catchment of the 186-km^2 Verzasca catchment. Above the gauges in Lavertezzo, they are both very little affected by human impact. Shortly after the gauge, the Verzasca flows into Lago di Vogorno, a retention lake for hydropower production. The elevation in the Pincascia catchment is 540 to 2500 m a.s.l. and in the Verzasca catchment 490 to 2900 m a.s.l. MeteoSwiss maintains a network of automatic stations that record meteorological data every 10 min. Additionally, Canton Ticino maintains an automatic network of pluviometers, recording precipitation data every 30 min. One of these cantonal stations is the only rain gauge within the Verzasca basin (Figure 1), but outside the Pincascia sub-basin. In addition to the automatic meteorological measurement station and the radar QPE (Germann *et al.*, 2006b), an ensemble of weather radar QPE is also generated (REAL, Germann *et al.*, 2009). This radar ensemble generator was implemented for research purposes by MeteoSwiss. Because the set-up is computationally very expensive, it is restricted to an area in Southern Switzerland (Figure 1). Within this area, the Verzasca is the only gauged catchment not disturbed by water management. Its unspoilt state

DOI: 10.1002/hyp

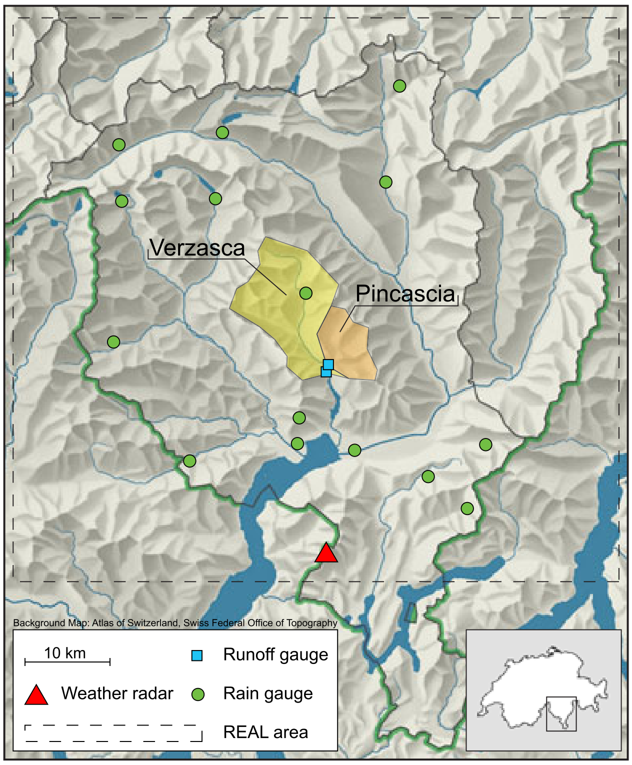

Figure 1. Southern Switzerland. Location of the test catchments Verzasca and Pincascia. The dashed square marks the region covered by the radar ensemble REAL. The symbols indicate the location of the weather radar (triangle), of the rain gauges (dots), and of the discharge gauges (squares)

combined with the good data availability and diversity makes the Verzasca very suitable for our study. The results are thus catchment-specific. Nonetheless, we can investigate the potential of the method, which is necessary to justify a potential future implementation on a bigger scale or in other regions prone to flash floods. The discharge measurements used for verification are available in hourly time steps. For the Verzasca, verified discharge observations are available from September 1989 to December 2010 and for the Pincascia from July 1992 to December 2010. All discharge data are provided by the Federal Office for the Environment (FOEN). Because our focus was on floods triggered by heavy thunderstorms, we excluded the months December to April from our analysis to avoid situations related to snow and snowmelt, and focused on the months May to November.

The Verzasca catchment is ideal for such experiments because the discharge time series is fairly long, high quality precipitation estimates are available, and the influence of human activities is small (e.g. Wöhling *et al.*, 2006; Germann *et al.*, 2009; Zappa *et al.*, 2011). Moreover, having Pincascia as a nested sub-catchment of the Verzasca catchment provides a rare opportunity to internally verify the probabilistic nowcasting system.

METHODOLOGY

Five different types of precipitation input data were used to force the semi-distributed rainfall-runoff model PREVAH (Viviroli *et al.*, 2009b). These include the deterministic radar QPE and precipitation fields interpolated from the automatic rain gauge data, as well as the radar ensemble and two parameter ensembles, resulting in five different continuous discharge nowcasts. The nowcasting range is 0 h, which means that the lead time of the discharge nowcast is given by the internal lead time of the catchment. The time period for which discharge nowcasts are generated and analysed is limited by the availability of the radar ensemble. Thus, continuous discharge nowcasts were generated from April 2007 to December 2010. Because we focused on flash floods, we restricted the evaluation of the time series to the months May to November because flash floods are unlikely to occur during the winter months.

All nowcast products were evaluated using different standard performance measures from atmospheric sciences and hydrology (Wilks, 2006). In addition to the evaluation of the 4 years' discharge nowcast time series, the performance of the different nowcast products for two events included in that time period was also analysed. A special feature of the analysis is the separate evaluation of the main tributary, the Verzasca basin, and of the sub-basin, Pincascia. Because the calibration of the hydrological model considered only discharge information from the Verzasca river gauge, all scores obtained for Pincascia are an independent internal verification of the data sets.

The hydrological model, the different types of precipitation input data, and the model calibrations and performance measures applied for the time series and event analyses are described in the following sections.

Hydrological model

The hydrological model used in this study is the semi-distributed rainfall-runoff model PREVAH (Gurtz *et al.*, 2003; Zappa *et al.*, 2003; Viviroli *et al.*, 2009b). The model's spatial discretization relies on the delineation of HRUs that take into consideration information on topography, land use, and soils (Gurtz *et al.*, 1999). PREVAH operates in this specific application on a spatial resolution of 500×500 m and is forced by hourly hydro-meteorological data. The meteorological input variables required to run the model are air temperature, water vapour pressure, global radiation, wind speed, sunshine duration, and precipitation (Viviroli *et al.*, 2009b). These variables are obtained from automatic meteorological surface stations. For each variable, a meteorological surface with a 500×500-m grid was interpolated with inverse distance weighting, whereas radar QPE were downscaled to meet the spatial properties of the hydrological model. For air temperature, water vapour pressure, global radiation, and wind speed, an elevation-dependent trend was considered (Zappa and Kan, 2007; Jaun and Ahrens, 2009; Viviroli *et al.*, 2009b). For this study, precipitation was estimated in five different ways (see subsequent sections). Depending on the air temperature, precipitation is treated as snow (Zappa *et al.*, 2003).

Precipitation data and ensembles

Precipitation was derived from several automatic stations in and around the catchments and from weather radar (Figure 1). These data were also used to create

DOI: 10.1002/hyp

ensembles of precipitation estimates. In this section, we introduce the different precipitation estimates used as input to the nowcasting system. Figure 2 summarizes the different nowcasting schemes adopted in this study and introduces the abbreviations for the different nowcasting schemes used.

Interpolated precipitation (PLU). The precipitation measurements from the automatic rain gauges in and around the catchments have a temporal resolution of 15 to 30 min but are aggregated to hourly values to meet the requirements of the model set-up. The observations are interpolated over the catchment areas on a 500×500-m grid using inverse distance weighting. A bias correction factor was estimated by calibration and applied to all interpolated values. Such bias correction is applied in order to minimize the total discharge volume error of the model as observed at a catchment outlet during the calibration period. The bias correction addresses several systematic errors in the modelling chain (undercatch of the pluviometers, spatial interpolation errors, suitability of the available gauge networks, and errors in the estimation of evapotranspiration). Thus, this correction factor is the most sensitive tunable parameter of PREVAH (Viviroli *et al*., 2009b).

Radar QPE (RAD). A second source of precipitation information was the weather radar located about 20 km south of the target areas (Figure 1). The radar QPE have a spatial resolution of 1 km^2 and a temporal resolution of 5 min. These were aggregated to hourly precipitation fields and downscaled to a 500×500-m grid.

Precipitation data from both rain gauges and from the weather radar are affected by uncertainty (Pappenberger *et al*., 2009; Krajewski *et al*., 2010). Precipitation data from rain gauges are very accurate only at the site of the rain gauge itself, and the interpolation conducted may not account for all the small-scale rainfall variability. This high spatial variability in precipitation plays, however, a major role in runoff generation in relatively small flash flood prone catchments. With weather radar, it is possible to estimate the precipitation with a high spatiotemporal resolution (Schiemann *et al*., 2011). To correct for several of the errors inherent in radar reflectivity measurements,

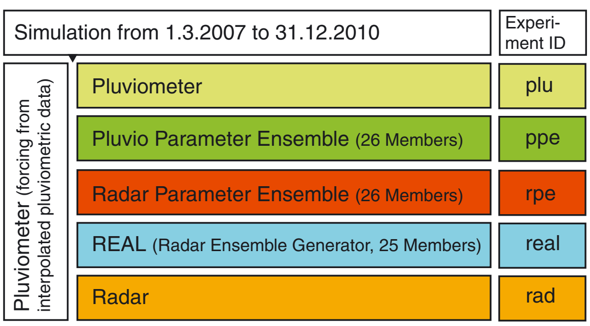

Figure 2. Experiment set-up for the different nowcast products. In the text as well as in graphs and tables, the different nowcast types are referred to by the experiment ID

MeteoSwiss developed and implemented a series of algorithms, which led to a significant improvement, but the residual uncertainty for applications in hydrology is still relatively high (Germann *et al*., 2006a). To account for this uncertainty, we therefore adopted the concept of ensembles.

Radar ensemble QPE (REAL). An elegant way to describe the residual uncertainty in radar QPE is to generate an ensemble of radar precipitation fields. For this purpose, Germann *et al*. (2009) developed a radar ensemble generator designed for usage in the Alps using LU decomposition (REAL). To produce the radar ensemble (REAL), the radar precipitation field is perturbed with correlated random noise. Thus, each member of the radar ensemble is then the sum of the deterministic radar QPE and a stochastic perturbation. Through this procedure, the residual space–time uncertainties in the radar estimates can be taken into account. The perturbation fields are generated by combining stochastic simulation techniques with detailed knowledge about the space–time variance and auto-covariance of radar errors (Germann *et al*., 2006a). REAL consists of 25 members and has been generated hourly in real time since 2007 on a spatial grid of 2×2 km^2 (Zappa *et al*., 2008). For a detailed description of the method applied to generate REAL, refer to Germann *et al*. (2009). To meet the requirements of the hydrological model, the radar ensemble was downscaled to a grid of 500×500 m.

Parameter ensembles (PPE and RPE). Besides the uncertainty associated with meteorological input, uncertainty arising from the model parameterization also has an influence on the resulting discharge simulation. The parameter set normally used at the initialization of the model simulations for the Verzasca and Pincascia originates from a default calibration used in previous applications (Wöhling *et al*., 2006; Ranzi *et al*., 2007). This calibration aims to find the parameter set that performs best in the simulation of the average flow and that has the smallest volume error between the observed and simulated time series (Zappa and Kan, 2007; Viviroli *et al*., 2009b). The deterministic nowcasts (PLU and RAD) were produced with this calibration.

As no model structure can perfectly represent the natural conditions and processes in a catchment, it is more honest to apply the concept of equifinality and look for an ensemble of equally likely behavioural parameter sets (Beven, 2006). A Monte Carlo (MC) experiment was therefore conducted to find equifinal model parameter sets. During this MC experiment, the seven most important parameters for conditioning the precipitation input and surface runoff generation were allowed to change randomly. To decide whether a model run is behavioural or not, a combination of two goodness-of-fit measures was chosen as an objective function (Zappa *et al*., 2011). These measures were the well-known Nash–Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) and the sum of weighted absolute errors (SWAE) proposed

DOI: 10.1002/hyp

in Lamb (1999) and Viviroli *et al.* (2009a). This likelihood function is tailored to fit peak discharge. A detailed description of this compound measure of performance can be found in Zappa *et al.* (2011). All MC runs were ranked according to this likelihood function. Finally, 26 parameter sets were chosen around the 95% ranking. In these 26 sets, the precipitation bias correction parameter (see above) ranged between +2% and +21%, and the threshold coefficient for triggering surface runoff (Gurtz *et al.*, 2003) ranged between 30 and 49 mm. The deterministic meteorological input with precipitation information from interpolated rain gauge data and radar QPE was then propagated through the hydrological model 26 times to generate ensemble hydrographs.

Performance measures

To assess the performance of the different schemes for ensemble discharge nowcasts, we considered several verification metrics. The *Brier skill score* (BSS) is among the most common skill scores used for the verification of probabilistic forecasts of dichotomous events, and it allows a direct comparison of deterministic and probabilistic predictions (Wilks, 2006). The BSS is based on the Brier score (BS). The BS is the mean squared error of the probability forecasts, given the observed outcome of an event (yes/no):

$$\mathrm{BS} = \frac{1}{n}\sum_{d=1}^{n}(y_d - o_d)^2, \qquad (1)$$

where o_d indicates whether the observation exceeded a predefined threshold (yes: $o_d = 1$; no: $o_d = 0$) and y_d is the forecast probability to exceed this predefined threshold ($0 \le y_d \le 1$).

The BSS indicates the improvement of the forecast in BS compared with a reference forecast (Wilks, 2006). In our study, the reference forecast is the climatological probability of exceedance of the predefined threshold, taking the months May to November of the whole observation period into account. A perfect forecast has a skill of 1, whereas forecasts worse than the climatological forecast have a skill below 0.

$$\mathrm{BSS} = 1 - \frac{\mathrm{BS}}{\mathrm{BS_{Cl}}} \qquad (2)$$

As high-intensity events are rather sparse, bootstrap confidence intervals were estimated for the BSS (Efron, 1992; Addor *et al*., 2011). 500 random samples of nowcast–observation pairs of daily maxima were drawn with replacement from the 856 days in the study period (May to November of the years 2007 to 2010). For each bootstrap sample, the BSS was computed, which meant that confidence bounds could then be estimated. See the box plots showing the distribution of the estimated BSS values in Figure 4.

The *false alarm ratio* (FAR) and the *probability of detection* (POD) are measures for deterministic predictions. For these, the ensembles were therefore reduced to their medians.

The FAR is the fraction of forecasted threshold exceedances, which in the end did not occur. The range of FAR is 0 to 1 and the best is 0.

$$\mathrm{FAR} = \frac{b}{a+b}, \text{where } a = \text{hit}, b = \text{false alarm} \qquad (3)$$

The POD is the proportion of correctly forecasted threshold exceedances to the number of times the event really happened.

$$\mathrm{POD} = \frac{a}{a+c}, \text{where } a = \text{hit}, c = \text{miss} \qquad (4)$$

The POD value ranges from 0 to 1, where 1 is the perfect score. Because the POD is only sensitive to missed events and not to false alarms, it can always be improved by predicting an event more often. However, this would directly result in a higher false alarm ratio and, for extreme events, in an over-forecasting bias (Bartholmes *et al*., 2009). It is therefore reasonable to show POD and FAR in combination.

Furthermore, we compute rank histograms (RHs). RHs show where the observation is ranked in relation to the entire ensemble, which means that they have one class more than the number of ensemble members. It provides information about the spread and bias of the REAL and parameter ensembles. BSS, FAR, and POD are calculated for five different exceedance thresholds, corresponding to the 0.6, 0.7, 0.8, 0.9, and 0.95 quantiles of the climatology (including May to November of the whole observation time series). The RHs are only computed for the 0.8 and 0.95 quantiles.

For all data types, the nowcasts of the years 2007 to 2010, which always included the months May to November, were analysed using NSE. Additionally, two selected events were analysed with NSE, which is a common efficiency measure in hydrology. It measures the relative improvement of the simulation compared to the mean of the observation (Schäfli and Gupta, 2007). The NSE thus penalizes errors in high flows more than errors in low flow conditions, which makes it most suitable for studies that focus on high discharge.

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{N}[q_{\mathrm{obs}}(t) - q_{\mathrm{sim}}(t)]^2}{\sum_{t=1}^{N}[q_{\mathrm{obs}}(t) - \bar{q}_{\mathrm{obs}}]^2} \qquad (5)$$

where $q_{\mathrm{obs}}(t)$ is the observed discharge at time step t, $q_{\mathrm{sim}}(t)$ is the simulated discharge, and $\bar{q}_{\mathrm{obs}}$ is the mean observed discharge over the whole simulation period of length N. For the probabilistic nowcasts, NSE values were calculated for the ensemble median.

RESULTS

In the first part of this section, we describe the main characteristics of the performance analysis for each

DOI: 10.1002/hyp

catchment. The compilation of all FAR and POD scores for the different thresholds, precipitation input, and catchments can be found in Tables I and II and are illustrated in Figure 3a and b. In the second part of this section, we present the analysis of the two selected events.

Verzasca

False alarm ratio and probability of detection. Values of FAR and POD are generally higher for radar-based nowcasts (RAD, RPE, and REAL) than for pluviometer-based nowcasts (PLU and PPE) (Figure 3a). FAR values rise between the 0.8 and 0.95 quantiles, most obviously for the radar-based nowcasts. The rise in POD between the 0.9 and 0.95 quantiles is, in contrast, more pronounced for the pluviometer-based products. RAD is associated with not only higher POD but also higher FAR values than PLU on all quantiles considered, which is in accordance with the differences in the probabilistic nowcasts PPE and RPE. PPE shows constantly lower POD and lower FAR than RPE. On the higher quantile, however, their POD values differ only very little (Figure 3a).

The deterministic nowcasts RAD and PLU behave similarly to their probabilistic complements RPE and PPE. The deterministic nowcasts show lower POD and lower FAR from the 0.6 to 0.8 quantiles than their probabilistic counterparts. This behaviour is more pronounced for PLU and PPE. For high quantiles (0.9 and 0.95), the deterministic RAD and PLU show higher POD and FAR than their probabilistic counterparts.

Table I. False alarm ratio values for the different thresholds (0.6–0.95) and the different precipitation inputs and catchments

FAR		**0.6**	**0.7**	**0.8**	**0.9**	**0.95**
Verzasca	PLU	0.08	0.04	0.03	0.11	0.17
	PPE	0.12	0.11	0.04	0.05	0.09
	RPE	0.20	0.19	0.14	0.18	0.35
	RAD	0.16	0.13	0.12	0.20	0.40
	REAL	0.15	0.13	0.13	0.21	0.42
Pincascia	PLU	0.07	0.01	0.02	0.00	0.00
	PPE	0.13	0.01	0.00	0.00	0.00
	RPE	0.18	0.05	0.04	0.02	0.06
	RAD	0.15	0.05	0.05	0.07	0.11
	REAL	0.13	0.05	0.05	0.05	0.10

Table II. Probability of detection values for the different thresholds (0.6–0.95) and the different precipitation inputs and catchments

POD		**0.6**	**0.7**	**0.8**	**0.9**	**0.95**
Verzasca	PLU	0.79	0.66	0.60	0.63	0.87
	PPE	0.86	0.74	0.65	0.60	0.82
	RPE	0.90	0.79	0.69	0.68	0.79
	RAD	0.86	0.76	0.66	0.72	0.90
	REAL	0.86	0.75	0.65	0.70	0.82
Pincascia	PLU	0.75	0.53	0.40	0.34	0.40
	PPE	0.82	0.60	0.40	0.30	0.27
	RPE	0.90	0.66	0.45	0.40	0.43
	RAD	0.85	0.57	0.44	0.46	0.51
	REAL	0.83	0.55	0.43	0.44	0.52

Brier skill score. Values of BSS are positive for all data types and thresholds (Figure 4a). PPE scores consistently higher BSS than the deterministic PLU, and RPE also seems to score slightly better BSS than the deterministic RAD, most clearly on the 0.95 quantile. PPE outperforms RPE, most prominently on the 0.95 threshold. There is not much difference in BSS between the two deterministic nowcasts, PLU and RAD, up to the 0.95 quantile, where PLU clearly outperforms RAD.

Both REAL and RPE outperform the deterministic RAD on the upper quantiles. BSS values for all radar-based nowcasts drop significantly between the 0.9 and 0.95 quantiles, whereas with PLU and PPE, BSS improves from the 0.9 to the 0.95 quantile. For the Verzasca, PPE and REAL outperform the other data types up to the 0.9 quantile. On the 0.95 quantile, both PPE and PLU are clearly better than REAL. The uncertainty for the BSS values is greater on the 0.95 quantile, and radar-based nowcasts especially have a large confidence interval, as obtained by bootstrapping the results 500 times (Figure 4a).

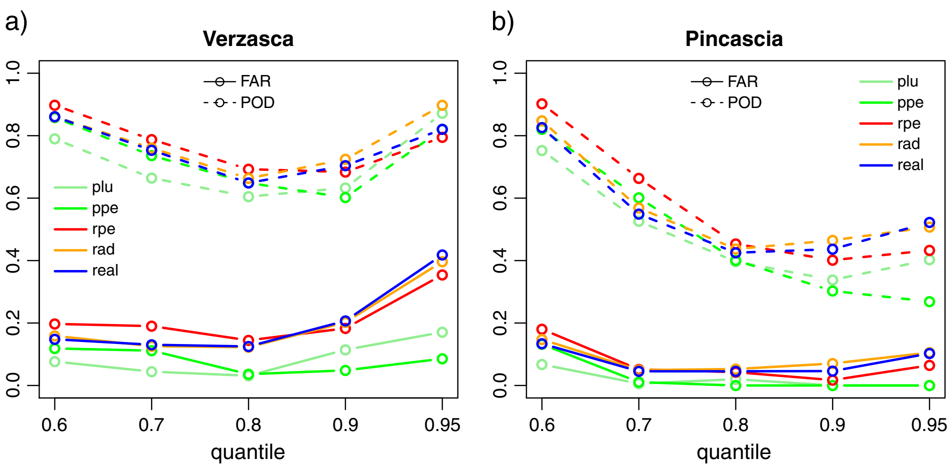

Figure 3. False alarm ratio (FAR) and probability of detection (POD) calculated on different threshold quantiles. The values for the ensemble nowcasts represent the results for their medians

DOI: 10.1002/hyp

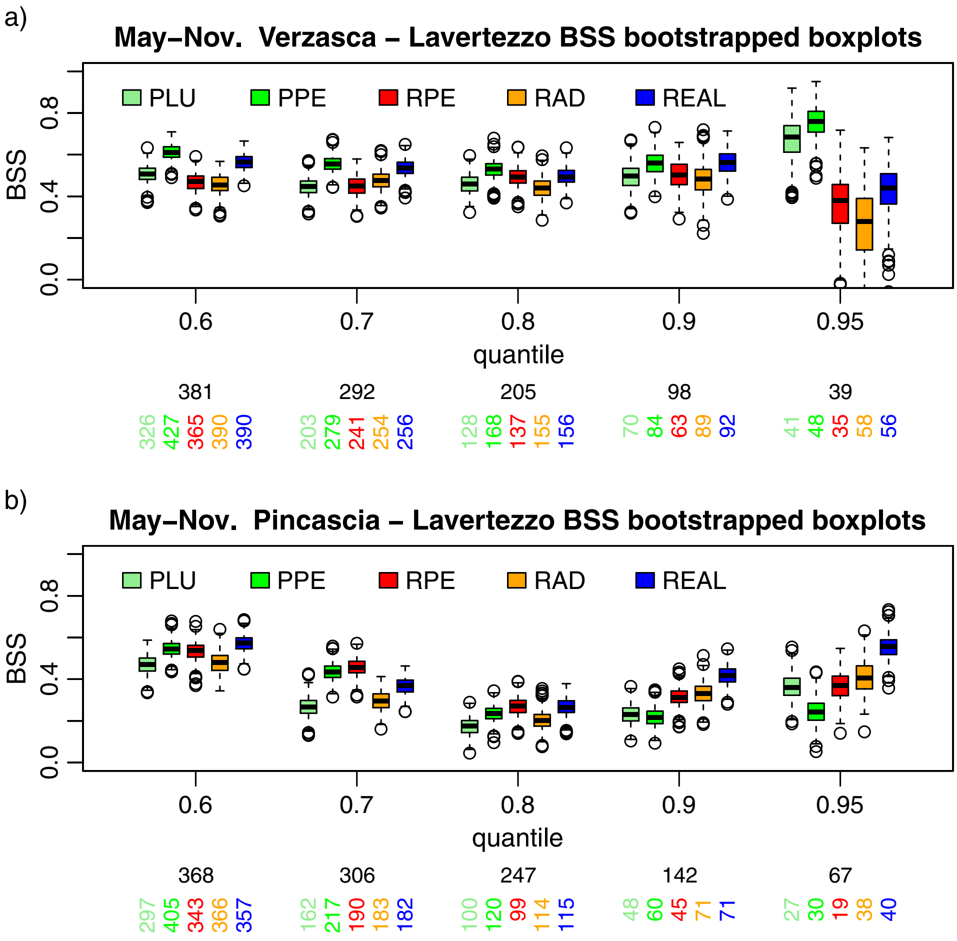

Figure 4. Brier skill scores (BSS) for different threshold quantiles for the years 2007 to 2010 (May to November of each year). The box plots describe the distribution of BSS derived from bootstrapping the sample. Horizontal numbers below the graphs indicate the number of days in which an exceedance of the threshold quantile was observed. Vertical numbers indicate for each nowcast type the number of days that were simulated to be above the threshold quantile. For ensembles, the number of ensemble members exceeding the threshold was divided by the size of the ensemble

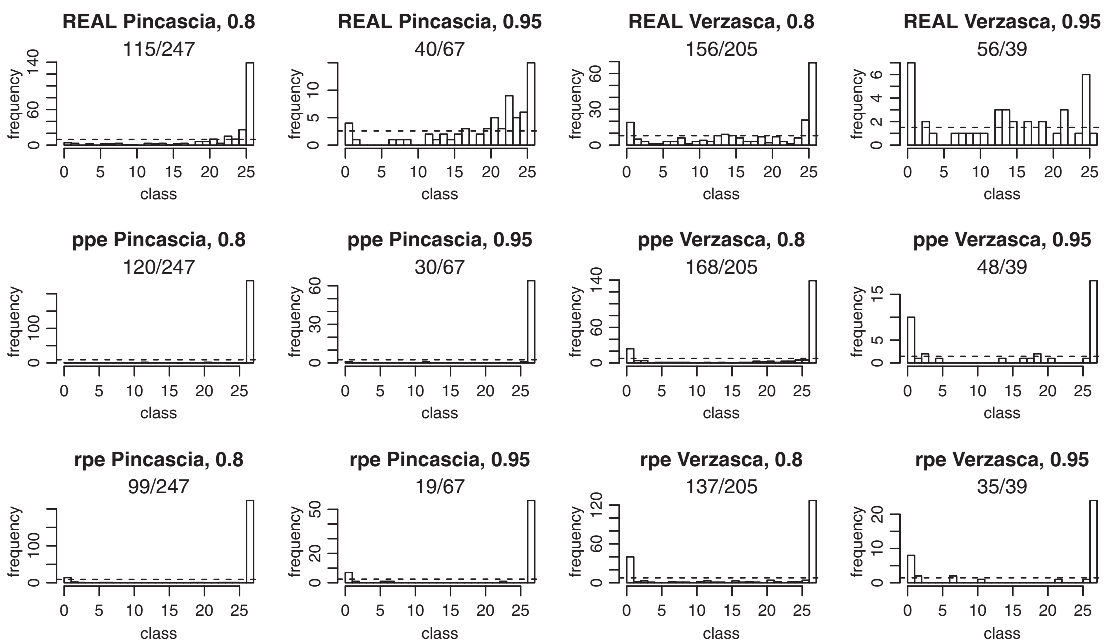

Figure 5. Rank histograms for the ensemble nowcasts for the 0.8 and 0.95 threshold quantiles. Numbers in the subtitle indicate in how many of the 856 days, included in the study period, the threshold quantile was exceeded by the ensemble and the observation, respectively. See Figure 2 for the different nowcast types

DOI: 10.1002/hyp

Rank histogram. The RHs on the 0.8 quantile are all J-shaped (Figure 5). Of the observations exceeding the climatological 0.8 quantile, 60% are above the PPE ensemble, 50% above the RPE ensemble, and 30% above the REAL ensemble. The ensembles are thus all underpredictive on the 0.8 quantile. On the 0.95 quantile, the RHs are U-shaped for all ensembles. The PPE and RPE, in particular, are underdispersive as the observation lies outside the ensemble in most cases. The REAL ensemble also shows a U-shape in the RH, but it is not as pronounced as for PPE and RPE. The 1st and the 25th class have most counts, but otherwise, the observations are almost uniformly distributed among the other classes.

Pincascia

False alarm ratio and probability of detection. On the 0.6 quantile, FAR and POD for the Pincascia nowcasts are about the same level as for the Verzasca ones, but for the higher quantiles, the values for Pincascia drop to a lower level (Tables I and II). For high quantiles (0.9 and 0.95), the radar-based nowcasts (RAD, RPE, and REAL) result in higher POD and FAR than the pluviometer-based nowcasts. On the lower quantiles, the POD for PLU and RAD do not differ much, but for the 0.9 and 0.95 quantiles, RAD clearly results in better POD values than PLU. The PPE and RPE show the same behaviour in POD values over the different quantiles, but the values for RPE are higher, especially on the high quantiles.

On high quantiles (0.9 and 0.95), PLU and RAD achieve higher POD values than the probabilistic PPE and RPE (Figure 3b).

Brier skill score. Values of BSS were positive for all data types and thresholds. Up to the 0.8 quantile, the BSS values for PPE were higher than for PLU, but for the highest quantiles, PLU outperformed PPE. The same behaviour can be seen between RPE and RAD, but the advantage of RAD above RPE on the high quantiles is not as clear as for PLU and PPE (Figure 4b).

The RAD nowcast outperforms PLU, and RPE outperforms PPE. This superiority of radar-based nowcasts over pluviometer-based nowcasts is most obvious for the upper quantiles (0.8–0.95). Considering only the radar-based nowcasts, REAL clearly outperforms the RAD and RPE on the two topmost quantiles. All data types show an improvement in BSS from the 0.9 to 0.95 quantile. The uncertainty in the estimated BSS values is only moderate and hardly changes over the threshold quantiles considered. Only on the highest threshold was there a slight increase in uncertainty.

Rank histogram. The RHs for PPE and RPE show a clear underpredicting bias on both the 0.8 and 0.95 quantiles. REAL is also underpredictive on both quantiles, but here, the underprediction is much smaller on the 0.95 quantile than on the 0.8 quantile (Figure 5).

Events

Two events, on 13 July 2008 and 17/18 July 2009, were chosen for an efficiency analysis of the different nowcast products. The events were not particularly extreme but have a return period lower than 2 years. The event of 17/18 July 2009 is ranked the fourth highest for the Verzasca catchment within the period from April 2007 to December 2010. The discharge reached 292.3 m^3/s in the Verzasca and 58.9 m^3/s in the Pincascia (Figure 6). The event of 13 July 2008 is ranked the tenth highest for the Verzasca catchment within the analysed period. Discharge reached 162.8 m^3/s in the Verzasca and 67.2 m^3/s in the Pincascia (Figure 7). Compared to the event in 2009, the Pincascia catchment was more affected by this event than the Verzasca catchment.

Visual evaluation of the 17/18 July 2009 event. For the Verzasca, the REAL-based ensemble runoff nowcast has the largest spread of all the ensembles. However, the REAL median shows the same characteristics as RAD and RPE. The timing and magnitude are quite good, but the volume of the event is overestimated. PLU and PPE result in a good timing of the rising limb and the peak, and the magnitude of the peak flow is also estimated well. The recession after the peak is too slow with PLU and PPE, resulting in a volume error on the falling limb. In the Pincascia, two peaks were observed for the event of 17/18 July 2009, the first smaller one (37.2 m^3/s) in the early afternoon and the main one (58.9 m^3/s) 11 h later, just after midnight. REAL has a very large spread especially around the first, smaller peak and overestimates the event, but REAL is the only nowcast that captures the small peak on 15 July. PPE underestimates the smaller peak on 17 July but estimates the magnitude and timing of the main peak quite well. The recession after the main peak is a bit too slow at the beginning of the falling limb. PLU underestimates the first peak and overestimates the second peak, but it estimates the timing of the peaks quite well. RAD overestimates the first peak by 10 m^3/s and keeps that level through to the second peak. The timing for both peaks is 1–2 h early. RPE has a good estimate of the first peak and rises up to the second peak, but it does not show the recession observed between the two peaks. This behaviour is also seen for the REAL median. The timing of the second peak is about 1 h early, but RPE has a good estimate of the falling limb.

Visual evaluation of the 13 July 2007 event. For this event, the REAL nowcast also has a big spread but shows all the minor and major peaks for this event, especially for the Pincascia (Figure 7).

For the Verzasca, PLU and PPE show the same behaviour, underestimating the event and only signalling the main peak 4 h late. RAD and RPE show a signal for both the minor and the main peak on 13 July 2008 and for the minor peak on 14 July. The timing is about 2 h early for the main peak and the magnitude a bit underestimated for 13 July, but for the minor peak on 14 July, RAD and RPE show an overestimation.

For the Pincascia, the radar-based nowcasts seem to generally represent the event better than the pluviometer-based nowcasts. The observation of the Pincascia lies most of the time within the range of the REAL nowcast. REAL

DOI: 10.1002/hyp

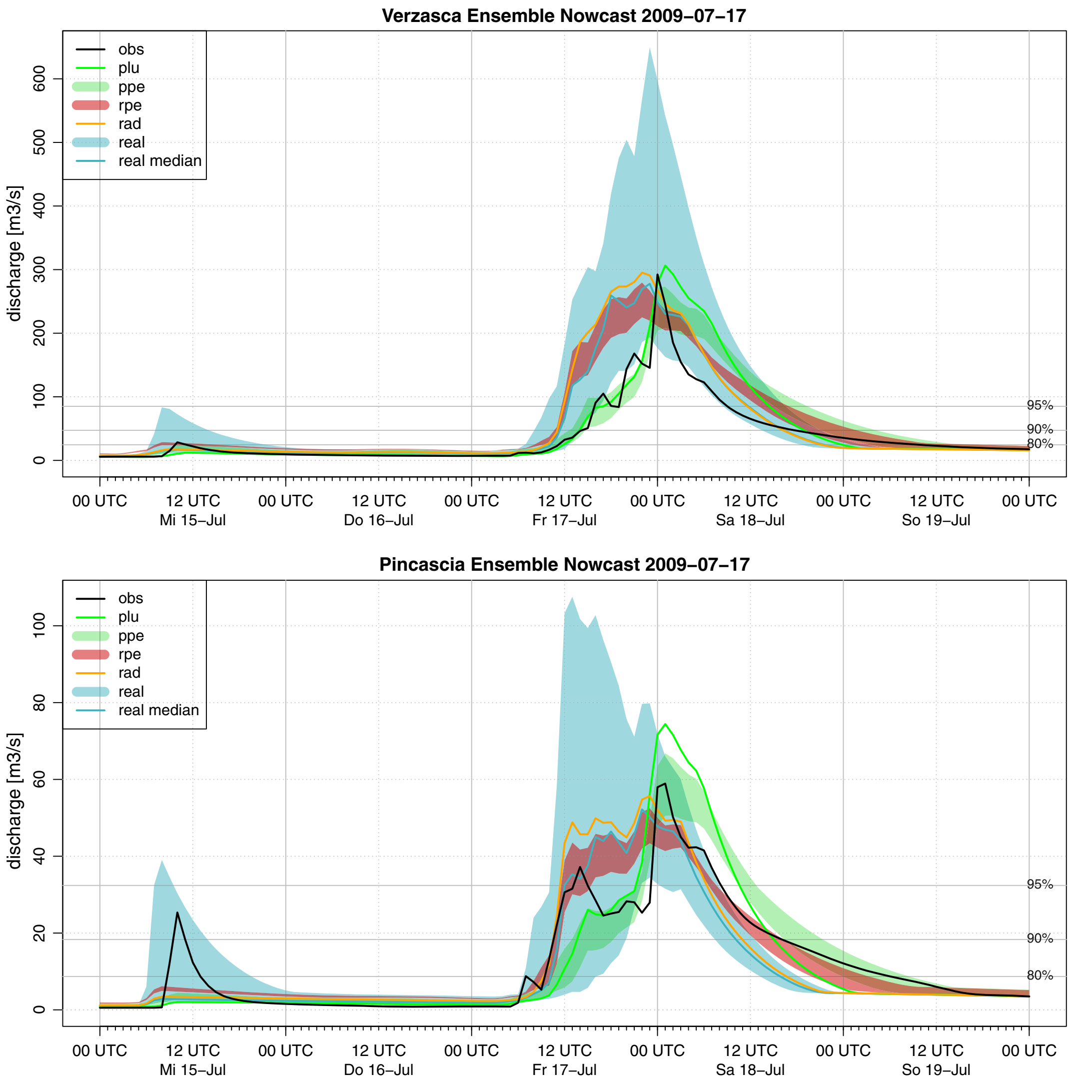

Figure 6. Continuous nowcast for the events on 17 July 2009

shows a signal for all the minor peaks and the main peak of the event. The timing of the whole event is a bit early for REAL. RAD and RPE represent the main characteristics of the event and show a signal for the three highest peaks, but they also underestimate the event. The timing of the main peak is about 1–2 h early. PLU and PPE underestimate the event. They show a signal only for the main peak and are 4 h too late.

Evaluation with Nash–Sutcliffe efficiency. For the Pincascia, the best NSE is with the radar-based parameter ensemble nowcasts (RPE and REAL). For the event on 17 July 2009, the NSE of RPE and RAD differ significantly, but this is not the case for the event of 13 July 2008, where they only differ by 0.01 (Table III). For 17 July 2009, all nowcast products have NSE values between 0.65 and 0.83, but the median of the ensemble nowcasts show higher NSE values than the deterministic PLU and RAD. For 13 July 2008, a significant difference in efficiency can be observed between radar-based and pluviometer-based nowcasts, with radar-based nowcasts producing clearly higher NSE values.

For the Verzasca, the PPE reaches an NSE of 0.7, the highest NSE value for the event of 17 July 2009, and clearly outperforms the other nowcast products. PLU scores second, whereas RPE and REAL reach about the same level of efficiency, 0.42 and 0.45, respectively. RAD achieves a very low NSE (0.14) for this event. For the event of 13 July 2008, however, RPE, RAD, and REAL reach about the same NSE values (0.71, 0.68, and 0.68) and clearly outperform the pluviometer-based PLU and PPE nowcasts (0.41 and 0.34).

For the entire study period, the NSE values are best for pluviometer-based nowcasts, especially for the Verzasca catchment, where NSE of PLU and PPE lie clearly above the values for radar-based nowcasts. For the Pincascia, this signal is only moderate.

DISCUSSION

Verzasca

The decrease of POD values towards the higher quantiles along with the FAR values staying at a low level implies that the number of missed events increases and thus the nowcast quality decreases. For the pluviometer-based nowcast, the quality seems to improve again for higher threshold quantiles, as the rise in FAR is fairly moderate in relation to

DOI: 10.1002/hyp

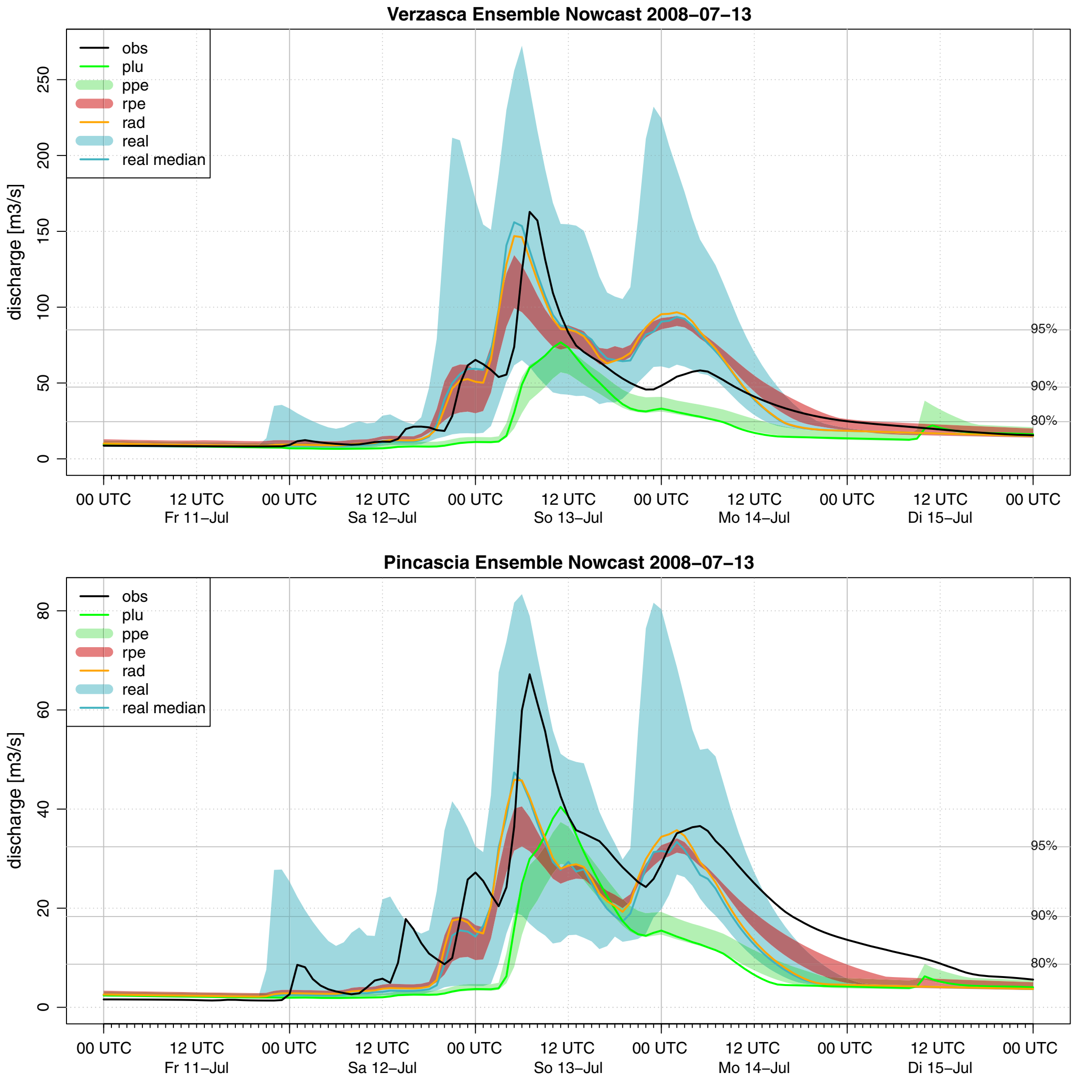

Figure 7. Continuous nowcast for the events on 13 July 2008

Table III. Nash–Sutcliffe efficiency (NSE) values for the two events shown in Figures 6 and 7, and for the whole time series 2007 to 2010, considering May to November of each year. For the ensembles, the NSE value of the ensemble median is shown

Catchment	Event	PLU	PPE	RPE	RAD	REAL
Verzasca	17.07.2009	0.59	0.70	0.45	0.14	0.42
	13.07.2008	0.41	0.34	0.71	0.68	0.68
2007–2010, May–November		0.81	0.83	0.65	0.58	0.60
Pincascia	17.07.2009	0.68	0.78	0.83	0.65	0.71
	13.07.2008	0.33	0.29	0.72	0.71	0.68
2007–2010, May–November		0.63	0.63	0.61	0.57	0.55

the rise in POD. This is also supported by the BSS values, which show an improvement of BSS for pluviometer-based nowcasts on high quantiles.

By comparing the deterministic PLU and RAD with their probabilistic counterparts PPE and RPE (BSS; Figure 4), a benefit can be observed with the probabilistic approach on most quantiles. In other words, including uncertainty results in better predictive skill. Furthermore, the pluviometer-based nowcasts clearly outperform the radar-based nowcasts on the highest quantiles. This reflects the fact that many uncertainties are involved in radar QPE in mountainous regions (Germann et al., 2006b). Moreover, PREVAH is calibrated on pluviometer data, and therefore, errors in the interpolated precipitation field are partly corrected. The results might look different if PREVAH were calibrated with weather radar data. It is not surprising that the probabilistic REAL and RPE, which try to account for these uncertainties, perform better than the corresponding

DOI: 10.1002/hyp

deterministic RAD nowcasts. The comparably good performance of the pluviometer-based nowcasts could also result from the fact that, for most high discharge events, the whole catchment is affected by large-scale heavy rainfall, and that under these conditions, even a single rain gauge inside the catchment can serve as a good reference for the situation in the whole catchment.

Pincascia

The predictability of discharge in the Pincascia is more challenging than in the Verzasca. POD and BSS values are generally on a lower level. The combination of low POD and low FAR indicates that positive predictions are reliable. In other words, a predicted event will most likely occur, but a high proportion of the events are missed and not predicted. In the small Pincascia catchment, errors in precipitation estimates and especially in their location have a big impact on the discharge prediction, and the rain gauge in the Verzasca catchment may not be affected by local storms that hit the Pincascia catchment. This explains the generally lower POD values for the Pincascia compared to the Verzasca (Table II).

A main difference between nowcasts for the Pincascia catchment and those for the Verzasca catchment can be seen in the performance of the different nowcast types, especially on the high quantiles. First, for the Pincascia catchment, radar-based nowcasts clearly outperform pluviometer-based nowcasts on the upper quantiles (Figure 4). Second, the parameter ensembles do not have an advantage over the deterministic nowcasts on high quantiles (Figures 4 and 3).

Thus, the fact that the deterministic nowcasts PLU and RAD outperform the parameter ensembles PPE and RPE in both POD and BSS shows that creating an ensemble to account for parameter uncertainty does not lead to an increase in skill under any condition, and especially not for extreme events in the Pincascia catchment, for reasons that are not entirely clear.

Looking only at the radar-based nowcasts, one can see that the parameter ensemble RPE shows less skill than the deterministic RAD, but the radar ensemble REAL reaches significantly higher scores. This is also due to the relatively high spread of REAL, which can also be seen in the RH (Figure 5).

Why the scores of RPE are so low is not clear. The good skill of REAL, however, indicates once again that, for small catchments like the Pincascia, the localization of the precipitation event plays the most important role in discharge predictions. REAL accounts for the space–time uncertainty in the radar QPE and therefore achieves relatively high BSS scores.

Event of 13 July 2008

The NSE values for the Pincascia achieved with the radar-based nowcasts were higher (0.68–0.72) than those with the pluviometer-based nowcasts (0.29–0.33), which corresponds with the visual evaluation of that event (Table III, Figure 7). This showed that radar-based nowcasts represent the characteristics of the event better than the pluviometer-based nowcasts do, which show a signal only for the main peak. This pattern is also observed for the Verzasca catchment, even though the NSE values for pluviometer-based nowcasts are slightly better than those for the Pincascia catchment. The main reason for the superiority of radar-based nowcasts for this event might be the location of the rainfall event. A query of the monthly weather report of July 2008 supports the interpretation that the upper part of the Verzasca catchment, where the only rain gauge is also located, was only slightly affected by the intensive downpours that took place during the nights of 12 and 13 July 2008, which led to local damage (MeteoSwiss, 2008). The Pincascia catchment, however, was in the sector of intensive rainfall. This explains why the Pincascia was comparatively more affected than the Verzasca and why the pluviometer-based nowcasts are only poor for this event.

Event of 17 July 2009

Hydrographs of PLU and PPE have the same shape for the Verzasca and the Pincascia for the event of 17 July 2009 (Figure 6). Because there is no rain gauge in the Pincascia catchment, the first, smaller discharge peak, which is only present for the Pincascia catchment, could not be detected. The visual impression that PLU and PPE describe the event in the Verzasca catchment better than the radar-based nowcasts is also reflected in the NSE values (Table III).

What is also noticeable for the Pincascia is that for the event in 2009, there is not much difference in the NSE values between the pluviometer-based and the radar-based nowcasts, whereas for the event in 2008, the radar-based nowcasts clearly achieve better NSE values (Table III). The reason for this difference has to do with the characteristics of the storm event. In 2008, the upper part of the Verzasca basin was not as much affected as the Pincascia catchment. At the rain gauge, located in the upper Verzasca catchment (Figure 1), 86.4 mm were measured on 12 July 2008 and 9.2 mm on 13 July 2008. During the storm in 2009, the whole Verzasca catchment was affected, and the rain gauge recorded a daily rainfall of 158 mm. Thus, in 2008 the radar nowcasts were clearly better, as the rain gauge was not representative of the situation in the whole catchment, in particular, not of the situation in the Pincascia sub-catchment. In 2009, the rain gauge did represent the situation in the catchment fairly well, and the interpolation with the other rain gauges resulted in a reasonable rainfall estimate for the whole catchment.

Looking at the Verzasca catchment, one will find that the most striking feature for the event on 17 July 2009 is the very low NSE value for RAD. Obviously, there must have been an additional source of uncertainty for the radar QPE. Most probably, this was due to the precipitation taking the form of hail. This assumption is supported by reports from Centro Meteo Lombardo, which recount that the heavy thunderstorms on 17 July 2009 included hail (CML, 2009). However, for both catchments, the radar-based ensemble

DOI: 10.1002/hyp

nowcasts (RPE and REAL) reach significantly higher NSE values than the deterministic RAD nowcasts (Table III). This is therefore an example where representing uncertainty in radar QPE by creating an ensemble pays off.

The efficiency analysis of the whole study period resulted in higher NSE values for pluviometer-based nowcasts. There are several reasons for this. The time period with low flow (dry period) is longer than that with high flow (wet period). This implies that, for most of the study period, the main source of uncertainty originates from the model parameterization and not from the meteorological input variable. Therefore, applying a parameter ensemble leads to better efficiency. As PREVAH is calibrated for the Verzasca catchment using interpolated rain gauge data, it is not surprising that the efficiency for pluviometer-based nowcasts is higher than for radar-based nowcasts and that efficiencies are generally higher for the Verzasca.

CONCLUSIONS

We have verified one of the first long time series of hydrological ensemble prediction systems (HEPS) applied to nowcasting in areas prone to flash floods. In comparison with the established use of HEPS (forcing by atmospheric EPS), we forced our hydrological model with weather radar ensemble QPE and with parameter ensembles of PREVAH. The comparison with the deterministic versions of the experiments allowed us to quantify the impact of uncertainties inherent in both the meteorological input and the model parameterization on the predictive skill of our HEPS. The evaluation included both the description of single events and the application of established verification metrics to the whole time series for two nested Alpine basins.

The evaluation of two events within the 4-year period analysed in this study shows that the performance of pluviometer-based and radar-based nowcasts depends to a great extent on the characteristics of the storm. However, the results for the evaluation of the whole study period show clear preferences in the applicability of the different nowcast products.

For the gauged and calibrated catchment Verzasca, pluviometer-based nowcasts are superior to radar-based nowcasts when dealing with flash flood simulations. We showed that their performance can be improved by accounting for parameterization uncertainty by adopting a parameter ensemble obtained by using the GLUE methodology.

For the uncalibrated and ungauged (with respect to precipitation) mountainous catchments, like the Pincascia, parameterization uncertainty plays a minor role compared to meteorological input uncertainty. We showed that, for this catchment, radar-based flash flood nowcasts are superior to nowcasts based on interpolated pluviometer data. The use of the radar ensemble REAL in particular proved to be significantly advantageous for nowcasting in this ungauged Alpine basins. This result shows the potential of REAL and is a prerequisite for the justification of a potential future implementation on a larger scale or in other poorly gauged regions prone to flash floods.

The various nowcast products discussed in this study could be used to generate initial conditions for subsequent forecast experiments with longer lead times. Initial tests with new radar extrapolation techniques (Mandapaka *et al.*, 2012; Panziera *et al.*, 2011) tailored to flash floods in mountainous regions are ongoing. The full potential of weather radar data for hydrological applications needs to be fully exploited. Long time series of weather radar data are needed to calibrate hydrological models and thus improve the predictive skill of radar-based discharge predictions.

ACKNOWLEDGEMENTS

This study was funded by the EU FP7 Project IMPRINTS (grant agreement no. 226555/FP7-ENV-2008-1-226555). MeteoSwiss is gratefully acknowledged for providing all the meteorological data used and FOEN for providing the observed discharge data. We would also like to thank the two reviewers for their helpful comments, Nans Addor for programming support and Silvia Dingwall for reviewing the language of this article.

REFERENCES

Addor N, Jaun S, Zappa M. 2011. An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): skill, case studies and scenarios. *Hydrology and Earth System Sciences* **15**: 2327–2347.

Barredo JI. 2007. Major flood disasters in Europe: 1950–2005. *Natural Hazards* **42**: 125–148.

Bartholmes JC, Thielen J, Ramos MH, Gentilini S. 2009. The european flood alert system EFAS – Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts. *Hydrology and Earth System Sciences* **13**: 141–153.

Berenguer M, Corral C, Sanchez-Diezma R, Sempere-Torres D. 2005. Hydrological validation of a radar-based nowcasting technique. *Journal of Hydrometeorology* **6**: 532–549.

Beven K. 2006. A manifesto for the equifinality thesis. *Journal of Hydrology* **320**: 18–36.

Beven K, Freer J. 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology* **249**: 11–29.

Beven KJ, Binley AM. 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes* **6**: 279–298.

Borga M, Anagnostou EN, Blöschl G, Creutin JD. 2011. Flash flood forecasting, warning and risk management: the HYDRATE project. *Environmental Science & Policy* **14**: 834–844.

Chiang Y-M, Hsu K-L, Chang F-J, Hong Y, Sorooshian S. 2007. Merging multiple precipitation sources for flash flood forecasting. *Journal of Hydrology* **340**: 183–196.

Choi HT, Beven K. 2007. Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework. *Journal of Hydrology* **332**: 316–336.

Cloke HL, Pappenberger F. 2009. Ensemble flood forecasting: A review. *Journal of Hydrology* **375**: 613–626.

CML, Centro Meteo Lombardo. 2009. 17 Luglio 2009: nubifragi e grandinate dall'Alta Pianura alle Prealpi. Retrieved November 1, 2011, from http://www.centrometeolombardo.com/content.asp?contentid=4307&ContentType=Archivio

Collier CG. 2007. Flash flood forecasting: What are the limits of predictability? *Quarterly Journal of the Royal Meteorological Society* **133**: 3–23.

Collier CG. 2009. On quality indicators for radar-based river flow forecasts. *Proceedings of the Institution of Civil Engineers-Water Management* **162**: 115–123.

Efron B. 1992. Jackknife-after-bootstrap standard errors and influence functions. *Journal of the Royal Statistical Society Series B-Methodological* **54**: 83–127.

Germann U, Berenguer M, Sempere-Torres D, Salvadè G. 2006a. Ensemble radar precipitation estimation — a new topic on the radar horizon. In *4th European Conference on Radar in Meteorology and Hydrology*. Barcelona; 559–562.

Germann U, Berenguer M, Sempere-Torres D, Zappa M. 2009. REAL - Ensemble radar precipitation estimation for hydrology in a mountainous region. *Quarterly Journal of the Royal Meteorological Society* **135**: 445–456.

Germann U, Galli G, Boscacci M, Bolliger M. 2006b. Radar precipitation measurement in a mountainous region. *Quarterly Journal of the Royal Meteorological Society* **132**: 1669–1692.

Gourley JJ, Giangrande SE, Hong Y, Flamig ZL, Schuur T, Vrugt JA. 2010. Impacts of Polarimetric Radar Observations on Hydrologic Simulation. *Journal of Hydrometeorology* **11**: 781–796.

Gurtz J, Baltensweiler A, Lang H. 1999. Spatially distributed hydrotope-based modelling of evapotranspiration and runoff in mountainous basins. *Hydrological Processes* **13**: 2751–2768.

Gurtz J, Zappa M, Jasper K, Lang H, Verbunt M, Badoux A, Vitvar T. 2003. A comparative study in modelling runoff and its components in two mountainous catchments. *Hydrological Processes* **17**: 297–311.

Hapuarachchi HAP, Wang QJ, Pagano TC. 2011. A review of advances in flash flood forecasting. *Hydrological Processes* **25**: 2771–2784.

He X, Refsgaard JC, Sonnenborg TO, Vejen F, Jensen KH. 2011. Statistical analysis of the impact of radar rainfall uncertainties on water resources modeling. *Water Resources Research* **47**: W09526.

Jaun S, Ahrens B. 2009. Evaluation of a probabilistic hydrometeorological forecast system. *Hydrology and Earth System Sciences* **13**: 1031–1043.

Krajewski WF, Villarini G, Smith JA. 2010. RADAR-RAINFALL UNCERTAINTIES Where are We after Thirty Years of Effort? *Bulletin of the American Meteorological Society* **91**: 87–94.

Krzysztofowicz R. 2001. The case for probabilistic forecasting in hydrology. *Journal of Hydrology* **249**: 2–9.

Lamb R. 1999. Calibration of a conceptual rainfall-runoff model for flood frequency estimation by continuous simulation. *Water Resources Research* **35**: 3103–3114.

Mandapaka PV, Germann U, Panziera L, Hering A. 2012. Can lagrangian extrapolation of radar fields be used for precipitation nowcasting over complex alpine orography? *Weather and Forecasting* **27**: 28–49.

Mandapaka PV, Villarini G, Seo B-C, Krajewski WF. 2010. Effect of radar-rainfall uncertainties on the spatial characterization of rainfall events. *Journal of Geophysical Research* **115**(D17110): 16. DOI: 10.1029/2009JD013366.

MeteoSwiss. 2008. *Witterungsbericht Juli 2008*. MeteoSwiss: Zürich.

Michelson D, Einfalt T, Holleman I, Gjertsen U, Friedrich K, Haase G, Lindskog M, Jurczyk A. 2005. Weather Radar Data Quality in Europe: Quality Control and Characterization In *COST 717 document*. Brussels; 87.

Molteni F, Buizza R, Palmer TN, Petroliagis T. 1996. The ECMWF ensemble prediction system: Methodology and validation. *Quarterly Journal of the Royal Meteorological Society* **122**: 73–119.

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models (1), a discussion of principles. *Journal of Hydrology* **10**: 282–290.

Panziera L, Germann U. 2010. The relation between airflow and orographic precipitation on the southern side of the Alps as revealed by weather radar. *Quarterly Journal of the Royal Meteorological Society* **136**: 222–238.

Panziera L, Germann U, Gabella M, Mandapaka PV. 2011. NORA–Nowcasting of orographic rainfall by means of analogues. *Quarterly Journal of the Royal Meteorological Society* **137**: 2106–2123.

Pappenberger F, Ghelli A, Buizza R, Bodis K. 2009. The skill of probabilistic precipitation prediction under observational uncertainties. *Journal of Hydrometeorology* **10**: 807–819.

Ranzi R, Zappa M, Bacchi B. 2007. Hydrological aspects of the Mesoscale Alpine Programme: Findings from field experiments and simulations. *Quarterly Journal of the Royal Meteorological Society* **133**: 867–880.

Rossa A, Bruen M, Fruehwald D, Macpherson B, Holleman I, Michelson D, Michaelides S. 2005. COST 717 Action - Use of Radar Observation in Hydrology and NWP Models. COST; 286.

Rossa A, Liechti K, Zappa M, Bruen M, Germann U, Haase G, Keil C, Krahe P. 2011. The COST 731 Action: A review on uncertainty propagation in advanced hydro-meteorological forecast systems. *Atmospheric Research* **100**: 150–167.

Rossa AM, Laudanna Del Guerra F, Borga M, Zanon F, Settin T, Leuenberger D. 2010. Radar-driven high-resolution hydro-meteorological forecasts of the 26 September 2007 Venice flash flood. *Journal of Hydrology* **394**: 230–244.

Schiemann R, Erdin R, Willi M, Frei C, Berenguer M, Sempere-Torres D. 2011. Geostatistical radar-raingauge combination with nonparametric correlograms: methodological considerations and application in Switzerland. *Hydrology and Earth System Sciences* **15**: 1515–1536.

Schäfli B, Gupta HV. 2007. Do Nash values have value? *Hydrological Processes* **21**: 2075–2080.

Smith RB. 1979. The Influence of Mountains on the Atmosphere. In *Advances in Geophysics*, Saltzman B (ed). Elsevier: New York; 87–230.

Villarini G, Krajewski WF, Ntelekos AA, Georgakakos KP, Smith JA. 2010. Towards probabilistic forecasting of flash floods: The combined effects of uncertainty in radar-rainfall and flash flood guidance. *Journal of Hydrology* **394**: 275–284.

Viviroli D, Mittelbach H, Gurtz J, Weingartner R. 2009a. Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland - Part II: Parameter regionalisation and flood estimation results. *Journal of Hydrology* **377**: 208–225.

Viviroli D, Zappa M, Gurtz J, Weingartner R. 2009b. An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools. *Environmental Modelling & Software* **24**: 1209–1222.

Werner M, Cranston M. 2009. Understanding the Value of Radar Rainfall Nowcasts in Flood Forecasting and Warning in Flashy Catchments. *Meteorological Applications* **16**: 41–55.

Wilks DS. 2006. *Statistical methods in the atmospheric sciences*. Elsevier: Amsterdam; 627.

Wilson JW, Crook NA, Mueller CK, Sun J, Dixon M. 1998. Nowcasting Thunderstorms: A Status Report. *Bulletin of the American Meteorological Society* **79**: 2079–2099.

Wöhling T, Lennartz F, Zappa M. 2006. Technical Note: Updating procedure for flood forecasting with conceptual HBV-type models. *Hydrology and Earth System Sciences* **10**: 783–788.

Yatheendradas S, Wagener T, Gupta H, Unkrich C, Goodrich D, Schaffner M, Stewart A. 2008. Understanding uncertainty in distributed flash flood forecasting for semiarid regions. *Water Resources Research* **44**: 17. DOI: 10.1029/2007WR005940.

Zappa M, Beven KJ, Bruen M, Cofino A, Kok K, Martin E, Nurmi P, Orfila B, Roulin E, Schröter K, Seed A, Stzurc J, Vehviläinen B, Germann U, Rossa A. 2010. Propagation of uncertainty from observing systems and NWP into hydrological models: COST-731 Working Group 2. *Atmospheric Science Letters* **11**: 83–91.

Zappa M, Jaun S, Germann U, Walser A, Fundel F. 2011. Superposition of three sources of uncertainties in operational flood forecasting chains. *Atmospheric Research* **100**: 246–262.

Zappa M, Kan C. 2007. Extreme heat and runoff extremes in the Swiss Alps. *Natural Hazards and Earth System Sciences* **7**: 375–389.

Zappa M, Pos F, Strasser U, Warmerdam P, Gurtz J. 2003. Seasonal water balance of an Alpine catchment as evaluated by different methods for spatially distributed snowmelt modelling. *Nordic Hydrology* **34**: 179–202.

Zappa M, Rotach MW, Arpagaus M, Dorninger M, Hegg C, Montani A, Ranzi R, Ament F, Germann U, Grossi G, Jaun S, Rossa A, Vogt S, Walser A, Wehrhan J, Wunram C. 2008. MAP D-PHASE: real-time demonstration of hydrological ensemble prediction systems. *Atmospheric Science Letters* **9**: 80–87.

IV

Flash-flood early warning using weather radar data: from nowcasting to forecasting

K. Liechti[1], L. Panziera[2], U. Germann[2] and M. Zappa[1]
[1]Swiss Federal Research Institute WSL, Birmensdorf, Switzerland
[2] MeteoSwiss, Locarno Monti, Switzerland

Abstract

This study explores the limits of radar-based forecasting for hydrological runoff prediction. Two novel probabilistic radar-based forecasting chains for flash-flood early warning are investigated in three catchments in the Southern Swiss Alps and set in relation to deterministic discharge forecast for the same catchments. The first probabilistic radar-based forecasting chain is driven by NORA (Nowcasting of Orographic Rainfall by means of Analogues), an analogue-based heuristic nowcasting system to predict orographic rainfall for the following eight hours. The second probabilistic forecasting system evaluated is REAL-C2, where the numerical weather prediction COSMO-2 is initialized with 25 different initial conditions derived from a four-day nowcast with the radar ensemble REAL. Additionally, three deterministic forecasting chains were analysed. The performance of these five flash-flood forecasting systems was analysed for 1389 hours between June 2007 and December 2010 for which NORA forecasts were issued, due to the presence of orographic forcing.

We found a clear preference for the probabilistic approach. Discharge forecasts perform better when forced by NORA rather than by a persistent radar QPE for lead times up to eight hours and for all discharge thresholds analysed. The best results were, however, obtained with the REAL-C2 forecasting chain, which was also remarkably skilful even with the highest thresholds. However, for regions where REAL cannot be produced, NORA might be an option for forecasting events triggered by orographic forcing.

1 INTRODUCTION

To issue early warnings about flash floods, information about the spatial and temporal distribution of precipitation is crucial. Catchments with steep slopes and shallow soils, which are typical in the Alps, react in particular very quickly to intense rainfall. Forecasting for flash flood events would thus help to extend the time available to issue warnings and implement safety measures. Producing such forecasts is, however, a very challenging task.

Hydrological forecasting has to deal with manifold problems. Not only is it very difficult to model the physical processes that affect runoff generation, but also the uncertainty about the distribution and intensity of the main triggering variable, the precipitation, poses significant challenges.

Precipitation measurements from rain gauges cover only small areas of a few square decimetres (Michelson, 2004; Sevruk, 1996), but they are then interpolated over tens or hundreds of square kilometres (Tobin et al., 2011; Velasco-Forero et al., 2009). Considering the very high spatial variability of precipitation, a problem of representativeness arises. It is already challenging enough to estimate precipitation distributions spatially when precipitation has occurred, but even more difficult to predict its spatial and temporal distribution in advance to be able to issue warnings and take preventive actions if needed to minimize any kind of loss.

The weather radar quantitative precipitation estimate (QPE) seems to be a very suitable product to detect the location of precipitation and to follow its development over time very closely because it is available at very high spatial and temporal resolutions. In Switzerland the information is provided every 5 minutes at a spatial resolution of 1 km (Germann et al., 2006). However, determining weather radar QPE is not an easy task, particularly in mountainous terrain, due to various sources of error, like beam shielding, ground clutter and hardware instabilities, amongst other (Germann et al., 2006; Szturc et al., 2008; Werner and Cranston, 2009). One approach to take these uncertainties into account is to use ensembles of weather radar QPEs (Germann et al., 2009; Liechti et al., 2012), but like rain-gauge data, radar QPEs are only available in realtime and not in advance.

A common way to predict precipitation is to use numerical weather prediction systems (NWP). They are run at different spatial and temporal resolutions, typically ranging from 2 to 10 km and from 24 to 240 hours of lead time (Montani et al., 2011; Zappa et al., 2008). One of the most detailed models available in Europe is the COSMO-2, which has a grid size of 2.2 km and 24 hours of lead time computed every 3 hours (Ament et al., 2011; Weusthoff et al., 2010).

These sources of precipitation estimates are all used as input in hydrological modelling. For flash flood early warning purposes, weather radar data is mainly used as input for nowcasts with zero lead time (Germann et al., 2009; Liechti et al., 2012; Zappa et al., 2011), which are then only meaningful within the response time of the modelled catchment, as Morin et al. (2009) describe. They developed and tested a flash-flood warning model for two catchments in the Dead Sea region based on real-time radar data. The system operates in both deterministic and probabilistic mode. For the probabilistic nowcasts they applied Monte Carlo simulations with an uncertainty range for both the radar QPEs and the model parameters. Despite the large amount of uncertainty they obtained acceptable model performance with their nowcasting system. For smaller catchments prone to flash floods, the response time of the catchment may be too short to issue useful warnings and to take mitigation actions in good time.

To give forecasts with a more useful lead time, methodologies based on Eulerian and Lagrangian persistence can be applied. Eulerian persistence keeps the current radar image frozen as a forecast for the near future (Germann and Zawadzki, 2002), while the Lagrangian persistence basically extrapolates the past motion of the precipitation into the future (Germann and Zawadzki, 2004; Mandapaka et al., 2012). Berenguer et al. (2005) did a hydrological verification of a radar-based nowcasting system by comparing stream-flow forecasts driven by S-PROG data (Seed, 2003) with forecasts driven by Eulerian and Lagrangian persistence. S-PROG is a simple extrapolation technique, based on Lagrangian persistence, that assumes a steady state for the motion of the rainfall field and also filters out the small-scale patterns of the rainfall field as the forecasting time increases. The verification of the system showed that an improvement in the precipitation forecast could be achieved with this method. However, the improvements in hydrograph prediction were not significantly better with S-PROG than with the simpler Lagrangian persistence.

To extend the lead time for flash-flood and flood early detection, several studies have also investigated the application of NWP forecasts in flash flood and flood early warning systems. Addor et al. (2011) compared flood forecasts driven by probabilistic and deterministic NWP forecasts. In their case study they found that, despite the coarser spatial resolution, the probabilistic forecast outperforms the deterministic forecasts for the whole forecast range of three days and also extends the lead time.

Similarly, Alfieri et al. (2012) analysed the performance of a NWP-driven flash-flood alert system. They used a 30-year meteorological reforecast (Fundel *et al.*, 2010) to derive warning thresholds from the hydrological model with the aim to be independent from any stream-flow observations. They calculated forecasts every third hour at a spatial resolution of 1 km with lead times up to 5 days and analysed their flash-flood forecasting system on the basis of a qualitative and quantitative performance analysis of the Verzasca catchment in southern Switzerland. The problems they encountered are well known: 1) only a limited amount of data is available for verification, which is why the warning thresholds are set very low to be able to do robust statistics, but these thresholds are then not really relevant for flash floods; 2) the catchment reacts very quickly to extreme precipitation and thus the interval at which the model operates is a limiting factor; and 3) NWP forecasts of convective precipitation events are not very accurate.

To address this last issue, Rossa et al. (2010) tested a hydro-meteorological forecasting chain that assimilates radar rainfall data into the NWP model COSMO-2 prior to processing the forecast data with a hydrological model. This allows the main convective systems to be introduced into the model state, which enhances the timing and localization of precipitation forecasts. This method seemed to improve discharge forecasts up to a lead time of three hours.

Up to now flash-flood early warning systems have either been run with NWP or, if run with weather radar data, they have been restricted to nowcasts with very limited lead time. Most of these studies, however, applied a deterministic approach.

Here we attempt a step from nowcasting to the radar-based forecasting of flash floods, evaluate two novel approaches to probabilistic radar-based flash-flood forecasting. The first is purely radar-based and provides forecasts for the next eight hours. It propagates analogue-based weather radar forecasts with a hydrological model and is designed for situations with orographic precipitation. The other approach combines a real-time radar ensemble nowcast (Germann et al., 2009) with the numerical weather prediction model COSMO-2. The resulting stream-flow forecasts are analysed and compared to deterministic radar-based and pluviometer-based forecasts. The aim of our study is to explore the space between radar-based nowcasting and radar-based forecasting and, in particular, to investigate the potential of purely radar-based flash-flood forecasting. Three basins of different sizes in the southern Swiss Alps were analysed, including the well-investigated Verzasca river basin (Alfieri et al., 2012; Germann et al., 2009)

2 THE HYDROLOGICAL MODEL

All the discharge forecasts in this study were produced with the semi-distributed rainfall-runoff model PREVAH (Gurtz et al., 2003; Viviroli et al., 2009). PREVAH operates at a spatial resolution of 500 m, however this grid is assembled to hydrological response units (HRU) containing information on land use, soil and topography (Gurtz *et al.*, 2003). The model is run at hourly intervals and is forced by spatially interpolated meteorological data. The meteorological variables required to run the model (air temperature, water vapour pressure, global radiation, sunshine duration, wind speed, and precipitation) are obtained from automatic meteorological ground stations and then interpolated with inverse distance weighting to form meteorological surfaces on a 500x500-m grid. Precipitation estimates from a weather radar and NWP data can also be used for model forcing (see e.g. Zappa et al., 2011). Prior to being used by PREVAH, the radar and NWP fields need to be downscaled to meet the resolution required by PREVAH. Due to the topographical variation in the catchments, an altitude-dependent gradient has to be considered for air temperature, wind speed, water vapour pressure and global radiation (Jaun and Ahrens, 2009; Viviroli et al., 2009; Zappa and Kan, 2007).

The adjustable parameters of the PREVAH model used in this study originate from a default calibration for the study areas obtained from previous applications (Ranzi et al., 2007; Wöhling et al., 2006). Data for the years 1992 to 2004 was used for model calibration and verification. The year 1992 was used as the initialisation period for the model, the years 1993 to 1996 for the calibration period and 1997 to 2004 for the verification period. The aim of the calibration is to find the parameter set that simulates the average flow best and that has the smallest volume error between the observed and simulated time series (Viviroli et al., 2009; Zappa and Kan, 2007). Discharge time series for verification were provided at hourly intervals by the Federal Office for the Environment (FOEN).

3 DATA

The precipitation nowcasts and forecasts used in our forecasting chains are described in the following sections. The methodologies we used have already been described in detail in previous publications. For details about the retrieval of weather radar and NWP products, see the articles cited below.

3.1 NORA – Nowcasting of Orographic Rainfall by means of Analogues

As precipitation in mountainous regions is influenced by orographic forcing, Panziera and Germann (2010) investigated the effects of orographic forcing on the rainfall patterns in the Lago Maggiore Region in Southern Switzerland (Fig. 3). They found strong relationships between the precipitation patterns and wind intensity, and the wind direction and air-mass stability present under orographic forcing. Based on this finding, they developed NORA (Nowcasting of Orographic Rainfall by means of Analogues), an analogue-based heuristic nowcasting system to predict orographic rainfall for the following eight hours (Panziera *et al.*, 2011). It involves finding earlier observations very similar to the current situation with respect to predictors describing the orographic forcing (four different mesoscale flows and air-mass stability) and two features of the radar rainfall field (fraction of rainy area and average rainfall). To speed up the process of finding analogues, all past weather radar data is reduced to an archive that only contains situations related to orographic forcing.

This archive was produced according to three different requirements: 1) the archive should be large enough to cover the whole range of the phenomena of interest; 2) it should be homogenous in terms of instrumental changes and data-processing techniques; and 3) the events selected should be long-lasting and widespread, as typically caused by large-scale supply of moisture towards the Alps. Isolated convection and air-mass thunderstorms were excluded from the archive. All these criteria finally resulted in an archive of 71 precipitation events observed between January 2004 and December 2009, corresponding in total to 3050 hours of rainfall.

To produce the NORA forecast, the historical situations most similar to the current one are searched for in the archive. This procedure is divided into two steps. In a first step, the 120 past instances most similar in terms of meteorological predictors (four mesoscale flows and air-mass stability) are chosen (forcing analogues). In a second step, the 12 analogues that, among the 120 forcing analogues, have the rainfall pattern most similar to the current one are picked. They constitute the final analogues. The NORA forecast is then produced according to the rainfall fields observed in the eight hours following each of the final analogues. This results in an ensemble of 12 members, one of which will, by construction, always be Eulerian persistence (Fig. 1). In this study, the number of forcing and final analogues of NORA was fixed, but in general it can be changed according to the archive size and the application. NORA is produced only if at least one of the four mesoscale winds can be estimated. Otherwise no orographic forcing is expected, and thus no NORA forecast is issued. The technical details about the algorithms behind NORA are given in Panziera et al. (2011).

NORA forecasts were originally issued in 5 minute time steps, but were aggregated to hourly time steps for our study to conform with the setting of the hydrological model. For the past events analysed in this study, the whole archive was searched for analogues. This meant that a hindcast of an event could also contain analogue situations that actually took place after the considered event in the past. Therefore, the 24 h following the initialisation of each NORA forecast were excluded from the archive in which the analogues were sought. Panziera et al. (2011) found that the results produced in this way did not differ significantly from results produced when only the hours of the archive preceding the NORA forecasts were included.

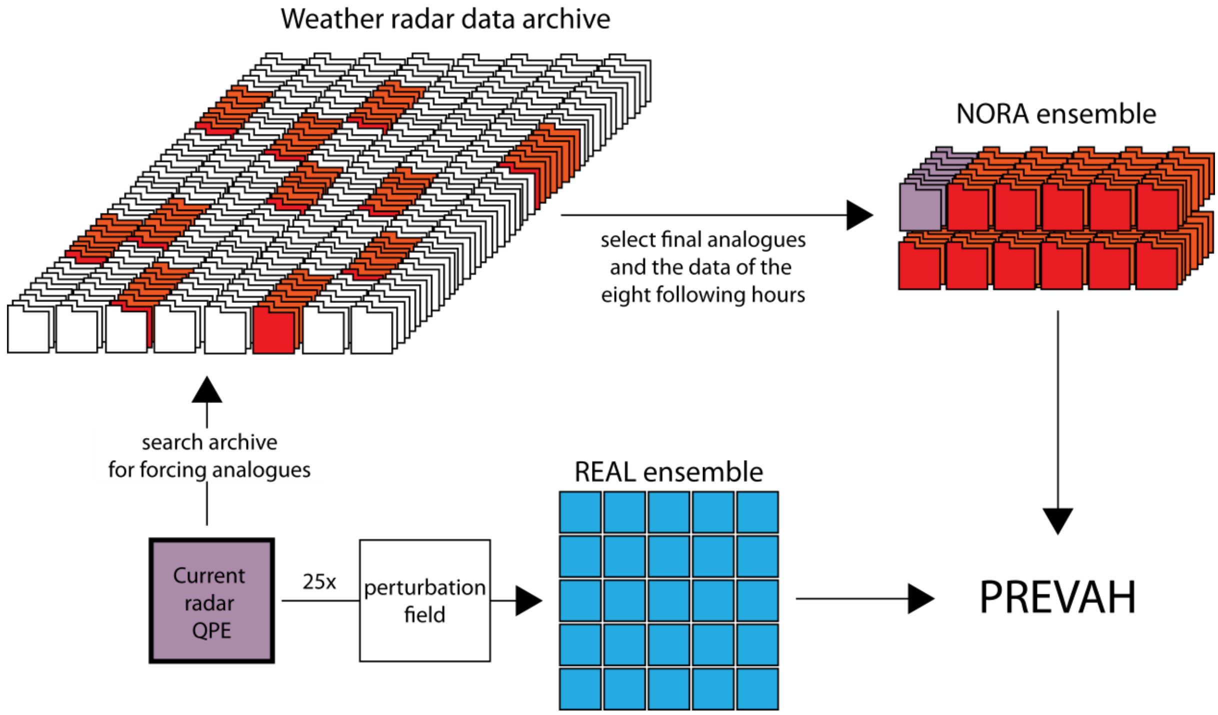

Fig. 1: Procedure to build the ensembles REAL and NORA. For the REAL ensemble, the current radar QPE is perturbed by a perturbation field 25 times to build an ensemble of 25 members. To build the NORA ensemble, a radar data archive is searched to find the situations most similar to the current radar QPE. Then those analogues and the data of the eight hours following each forcing analogue are extracted from the archive, and an ensemble of 12 members with 8 hours lead time each is built.

3.2 REAL – Radar Ensemble

REAL (Radar Ensemble generator designed for the Alps using LU decomposition) was developed by MeteoSwiss as a probabilistic real-time radar nowcasting tool (zero lead time). It provides an ensemble of 25 members, each of which results from the sum of the current radar image and a stochastic perturbation field (Fig. 1). This perturbation field is a combination of stochastic simulation techniques and detailed knowledge about the space-time variance and auto-covariance of radar errors (Germann *et al.*, 2009). To obtain this knowledge, a suitable network of meteorological ground stations is required. With this methodology the residual space-time uncertainties of the radar precipitation estimates are accounted for. REAL has been produced since May 2007 at hourly intervals with a spatial resolution of 2x2 km (Germann *et al.*, 2009) for the Lago Maggiore region in the Southern Swiss Alps (Fig. 3).

Coupling of REAL and COSMO-2

For our study we coupled COSMO-2 forecasts to the radar-ensemble nowcasts of REAL. COSMO-2 (C2) is a deterministic numerical weather prediction (NWP) model of the Consortium for Small-scale Modelling (COSMO). It has a lead time of 24 hours, a spatial resolution of 2.2 km and has been issued every three hours (0, 3, 6, 9, 12, 15, 18, 21 UTC) since the beginning of demonstration period of MAP D-PHASE (Rotach et al., 2009) in June 2007. The coupling with REAL implies that COSMO-2 meteorological input is actually propagated through the hydrological model every hour with 25 different initial conditions stemming from the nowcast obtained by forcing PREVAH with REAL.

3.3 Deterministic forecasts

In addition to the two probabilistic forecast chains with NORA and REAL-COSMO-2, we also tested the performance of two deterministic model chains. They are constructed like the REAL-C2 forecasts, unlike with REAL, the initial conditions are derived by driving PREVAH with interpolated rain-gauge data (PLUVIO) or the deterministic weather radar QPE (RADAR).

The data for the interpolated precipitation surfaces originated from automated rain-gauge stations, which have a temporal resolution of 15 to 30 minutes. These were aggregated to hourly values and interpolated with inverse distance weighting over the areas of the test catchments on a 500 x 500-m grid. Additionally, a bias correction factor was determined by calibration (Zappa and Kan, 2007) and applied to all interpolated values, in order to minimize the total discharge volume error at the catchment outlets (Viviroli *et al.*, 2009). The radar QPE was taken from the weather radar on Monte Lema (Fig. 3). They are available at a temporal resolution of five minutes and at a spatial resolution of 1 km^2, but were aggregated to hourly time steps and downscaled to a 500 x 500-m grid.

After the initialisation it takes 2.5 hours to assimilate, compute and disseminate COSMO-2. Since COSMO-2 is produced every three hours, this means that the COSMO-2 forecast is three to five hours old by the time it can be used for the hydrological forecast. Table 1 shows the schedule for coupling COSMO-2 forecasts to nowcasts forced by RADAR, PLUVIO or REAL.

Table 1: COSMO-2 forecasts coupled to discharge nowcasts forced by REAL, deterministic radar QPE (RAD) and interpolated rain-gauge data (PLU). Times are in hours UTC.

Initialisation of COSMO-2 forecast	***Available at***	***Start of discharge forecast***
00	02:30	03,04,05
03	05:30	06,07,08
06	08:30	09,10,11
09	11:30	12,13,14
12	14:30	15,16,17
15	17:30	18,19,20
18	20:30	21,22,23
21	23:30	00,01,02

3.4 Study period

The beginning of the study period was set to June 2007 according to the availability of COSMO-2. December 2010 defines the end of our study period. Due to the replacement of the weather radar on Monte Lema (Fig. 3), the continuous and homogeneous series of high quality weather radar data ends in early summer 2011. For the period June 2007 to December 2010, NORA forecasts were initialized on 1389 hours, when orographic precipitation occurred. These 1389 hours were distributed over 40 events. We analysed all 1389 forecasts, each of which consists of eight hours, for all forecasting chains included in our study. The 40 individual events are plotted sequentially in Fig. 2 for the Verzasca catchment, as an example, along with the NORA and REAL-C2 forecasts for 3 and 6 hours lead time.

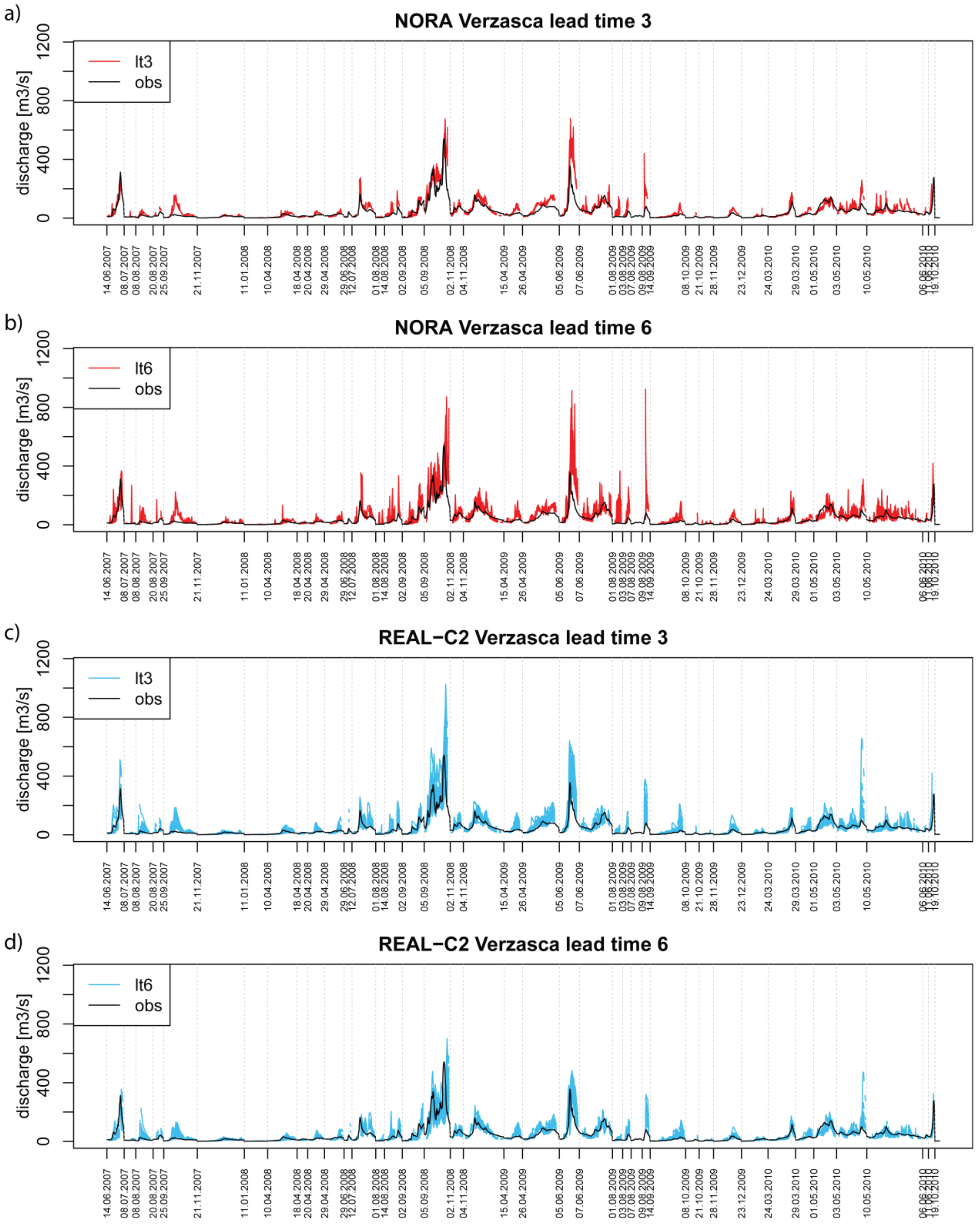

Fig. 2: NORA and REAL-C2 discharge ensemble for the Verzasca river for all 40 events in the study period. The panels a) and c) show the discharge ensembles at 3 hours lead time, and the panels b) and d) show the discharge ensembles at 6 hours lead time. The individual events are separated by dashed vertical lines. The dates given in the x-axis refer to the date of the beginning of each event.

4 The Catchments

Catchments were selected in the Lago Maggiore region in Southern Switzerland, where NORA and REAL are available. Until today these two products have been specially produced for research purposes and are therefore only available for this limited region (Fig. 3). In many catchments in the region, water is intensively managed for hydropower production. We therefore selected two smaller catchments which are not, or only slightly, affected by water management, as well as a large catchment to explore the effects of scale.

The Calancasca catchment is 120 km^2 and the smallest of the three catchments. The Calancasca valley is a subcatchment of the Ticino catchment, and is very rural and mountainous with steep slopes, ranges from 740 m a.s.l. to 3200 m a.s.l. in altitude. At the top of the catchment there is a small glacier, covering 1.1% of the catchment area. The catchment is little affected by hydropower, but some of the headwater is partly redirected to a hydropower plant in the neighbouring catchment to the east. This diversion is taken into account in the hydrological model with the routing module. Downstream of the Calancasca gauge, the stream water is stored in a small retention lake for hydropower production.

The Verzasca catchment is 186 km^2 in area ranging from 490 to 2900 m a.s.l.. It is very little influenced by human activity. At altitudes above the discharge gauge in Lavertezzo it is not affected by any water management but below the gauge, the river Verzasca flows into Lago di Vogorno, a retention lake for hydropower production. The basin is the main focus area for our research group. Wöhling et al. (2006) presented the results of model calibration and introduced an assimilation procedure aimed at improving the quality of initial conditions prior to and during an event. Germann et al. (2009) and Liechti et al. (2012) focused on the verification of the use of REAL as a forcing for real-time nowcasts. Zappa et al. (2011) developed and tested a methodology to quantify the relative contribution of different sources of uncertainty (forcing, initial conditions and model parameter estimation) to the total uncertainty of a real-time flood forecast.

The Ticino catchment is 1515 km^2 in area. It is much more densely populated and thus more influenced by human activity than the two small catchments. The main valley of the Ticino catchment is part of one of the main transit routes that cross the Alps. Hence the lower area of the catchment, where the valley is broad enough, is intensively used for industry and agriculture, whereas the steep slopes are only little used. Altitudes range from 220 m to 3400 m a.s.l. The influence of water management is substantial, but all water remains in the catchment and reaches the gauge in Bellinzona.

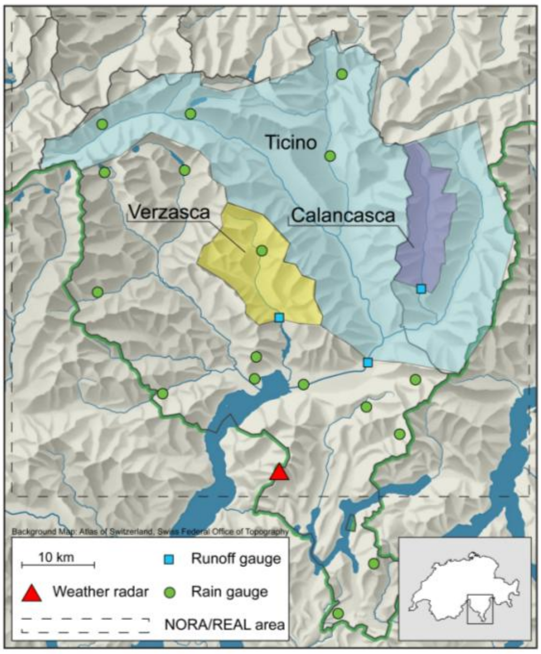

Fig. 3: Lago Maggiore region, Southern Switzerland, with test catchments, meteorological and hydrological stations and weather radar used in this study. The rectangle with dashed lines shows the area for which NORA and REAL have been produced.

5 Methodology

5.1 Experimental set up

Our experimental set up in hindcast mode for the five different forecasting chains consisted of a nowcasting part with zero lead time (realtime) and a forecasting part (Fig. 4). The nowcasting part was initialised five days prior to the onset of the NORA forecast (t0) by the model state derived from a reference run forced by pluviometer data (Fig. 4). This realtime part was run for four days, which meant the influences of the initial conditions are reduced at the start of the forecasting part at time t0. The five forecasting chains analysed are:

1) *NORA:* NORA forecast initialized by a deterministic RADAR nowcast.
2) *PERS:* the persistence of the current radar QPE at time t0 (i.e. taking the signal of t0 for the next eight hours) initialized by a deterministic RADAR nowcast.
3) *REAL-C2:* COSMO-2 forecast initialized by a probabilistic REAL nowcast.
4) *RAD-C2:* COSMO-2 initialized by a deterministic RADAR nowcast.
5) *PLU-C2:* COSMO-2 initialized by a deterministic PLUVIO nowcast.

Thus we were able to compare the performance of NORA with the performance of COSMO-2 given different initial conditions derived from discharge nowcasts of PREVAH forced by REAL, PLUVIO and RADAR. The diagram in Fig. 4 visually explains the different model chains and introduces the names and the colour scheme used from now on for the different forecasting chains.

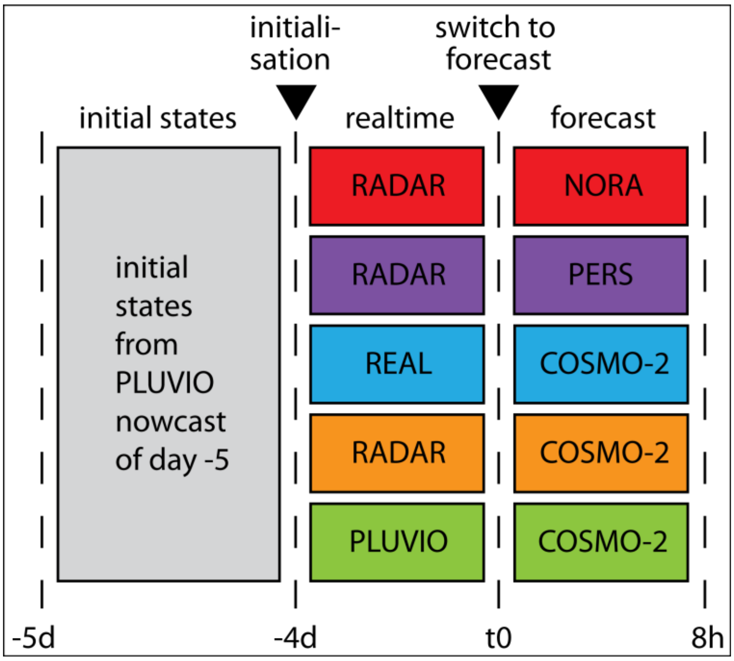

Fig. 4: Diagram showing the different forecasting chains.

5.2 Verification methods

The main objective of our study was to investigate a possible added value of NORA for flash-flood early warning. As NORA is limited to a lead time of eight hours, we concentrated our verification on these eight hours. We analysed the performance of the different forecasting chains for each lead time (1h-8h) separately, as well as for six different thresholds with the following measures of skill:

The *Brier Skill Score* (BSS) is an ideal measure to compare the performance of probabilistic and deterministic forecasts (Wilks, 2006). The BSS is based on the Brier Score (BS), which describes the quality of the forecast system in predicting the probability to exceed a predefined threshold by measuring the squared probability error. A perfect forecast system would have a BS of zero. In order to compare the different forecast systems to each other, we made use of the BSS, which sets the skill of the different forecasts in relation to a reference forecast. A perfect forecast has a BSS of 1, whereas forecasts worse than the reference forecast have a skill below 0. In our study, the reference forecast was the probability of exceedance for the predefined thresholds based on the sample climatology. The sample incorporated all discharge observations from hours covered by one or more NORA forecasts. This resulted in a sample size of 1788 hours. The thresholds analysed in our study correspond to the 0.5, 0.6, 0.7, 0.8, 0.9 and 0.95 quantile of the sample climatology, which we refer to as q50, q60, q70, q80, q90 and q95. As the sample is restricted to the hours covered by NORA, the actual values of the thresholds quantiles are higher than the ones used in our previous study (Liechti et al., 2012).

To estimate the uncertainty of the BSS values, we applied the bootstrapping method (Efron, 1992). Thus 500 random samples of forecast-observation pairs were drawn with replacement from the 1389 hours belonging to each lead time. The confidence limits (95%) shown in Fig. 6 were estimated by resampling the data with replacements for 500 times.

The *False Alarm Ratio (FAR)* and *Probability Of Detection* (POD) are interlinked and therefore shown together. Both are measures to evaluate deterministic predictions, where the ensembles were reduced to their medians. FAR is the fraction of the forecast threshold exceedances that turn out to be wrong. The best FAR value is zero, which means that each positive forecast was followed by a threshold exceedance. POD is the ratio of correctly forecast threshold exceedances to the number of times the event really happened. The best POD value is one, which means that each observed threshold exceedance was forecast. The POD is only sensitive to missed events and not to false alarms, and thus can always be improved by forecasting an event more frequently. This would, however, lead directly to an increase in false alarms and would, for extreme events, result in an overforecasting bias (Bartholmes et al., 2009; Wilks, 2006).

The *ROC area* (ROCA) is the area under the ROC (relative operating characteristic) curve. A ROC curve is drawn in a ROC diagram, which incorporates information on the POD (y-axis) and the false alarm rate (x-axis) for the whole range of forecast probabilities. The false alarm rate is the fraction of non-occurrences for which a threshold exceedance was forecast. A perfect forecast will result in a ROC curve connecting the points (0/0), (0,1) and (1/1) of the ROC diagram. An unskilful forecast will not lie above the diagonal (0/0),(1,1). Thus the area under the ROC curve is a convenient way to express the degree of discrimination. ROC is not, however, sensitive to an overall bias, which means that ROC actually indicates the potential skill, that would be achieved if the forecasts were correctly calibrated (Wilks, 2006). Therefore we also show the bias of the different forecasting chains.

6 RESULTS

First we show how the spread of the two ensembles NORA and REAL-C2 generally develops over lead time. We then present the results for the three catchments separately. The results of the analysis with ROC area are summarized for all catchments together. Finally, we present a forecast for the Calancasca as it appears in operational mode.

6.1 Chained time series

In Fig. 2 all events in the study period are plotted sequentially together with the forecasts with 3 and 6 hours lead time. The spread of the NORA ensemble increases with the lead time for all catchments.

However the spread of the REAL ensemble behaves differently in the Verzasca catchment than in the Ticino and Calancasca catchments. In Ticino and Calancasca the spread of the REAL ensemble stays about constant over the eight hours analysed (not shown), but in the Verzasca catchment the spread of REAL decreases with longer lead times. This is possibly due to the nature of the events included in the study period and is further discussed in section 7.2. For the Ticino and Calancasca catchments the spread of REAL is larger than that of NORA for all lead times, although for Calancasca the difference is relatively small from 6 hours lead time on. For Verzasca, the spread of REAL is only larger than that of NORA for up to 4 hours lead time, and from 6 hours lead time onwards NORA forecasts have a larger spread than REAL forecasts.

6.2 Calancasca

BSS values for REAL-C2 generally decrease with increasing threshold and longer lead times. REAL-C2 shows skill on all thresholds and all lead times. The highest BSS values are reached with q60 (0.56-0.6), but for q90 and q95 BSS values are still as high as 0.35 to 0.4 (Fig. 6b). NORA shows lower scores than REAL-C2, and its BSS values range between 0.35 to 0.4 for q50 to q80. For the highest thresholds BSS values are lower, while for q90 and q95 they range between 0.15 and 0.25. BSS values for PERS clearly decrease with lead time (Fig. 6a). The highest score is reached at q60. BSS values for q90 and q95 are below 0.2, while q90 shows no skill for lead time 8 (Fig. 6a) and q95 shows no skill after 3 hours lead time (Fig. 6b). RAD-C2 also shows skill on all thresholds and lead times, but this decreases with lead time (Fig. 6a). RAD-C2 forecasts reach BSS values between 0.3 and 0.4 for q60 to q80 and lead times up to 6 hours. The performance on q70 and q80, however, fall below 0.3. For q90 and q95 BSS values are generally lower than 0.2 (Fig. 6b). BSS values for PLU-C2 are highest on q70 and are above 0.6 up to 6 hours lead time. Additionally PLU-C2 outperforms all other forecast chains on this threshold. For the highest quantile PLU-C2 also shows skill over all lead times, varying between 0.14 and 0.36.

The probability of detection (POD) for PLU-C2 is higher than for the other forecast chains, as are the FAR values for thresholds q80 to q95 (Fig. 5). POD and FAR for PLU-C2 behave symmetrically from q70 to q95, which is not the case for the other forecast chains. POD for REAL-C2, NORA, PERS and RAD-C2 rapidly decrease above q60. FAR are lowest for REAL-C2 on all quantiles except q95. FAR and POD for NORA, PERS and RAD-C2 are about the same. FAR values range between 0.1 and 0.3 and POD values drop from 0.9 at q60 to 0.2-0.3 at q95. If we increase the lead time from 3 to 6 hours, the main difference is with the q95 threshold, where the FAR values are highest for all forecasting chains (Fig. 5). The different behaviour of the different forecasting chains is also mirrored in the bias. Forecasts for Calancasca have an underforecasting bias above q60 for all radar-based forecasts. This is most pronounced for REAL-C2. PLU-C2 performs best and is hardly biased above q60. This behaviour does not change with increasing lead time.

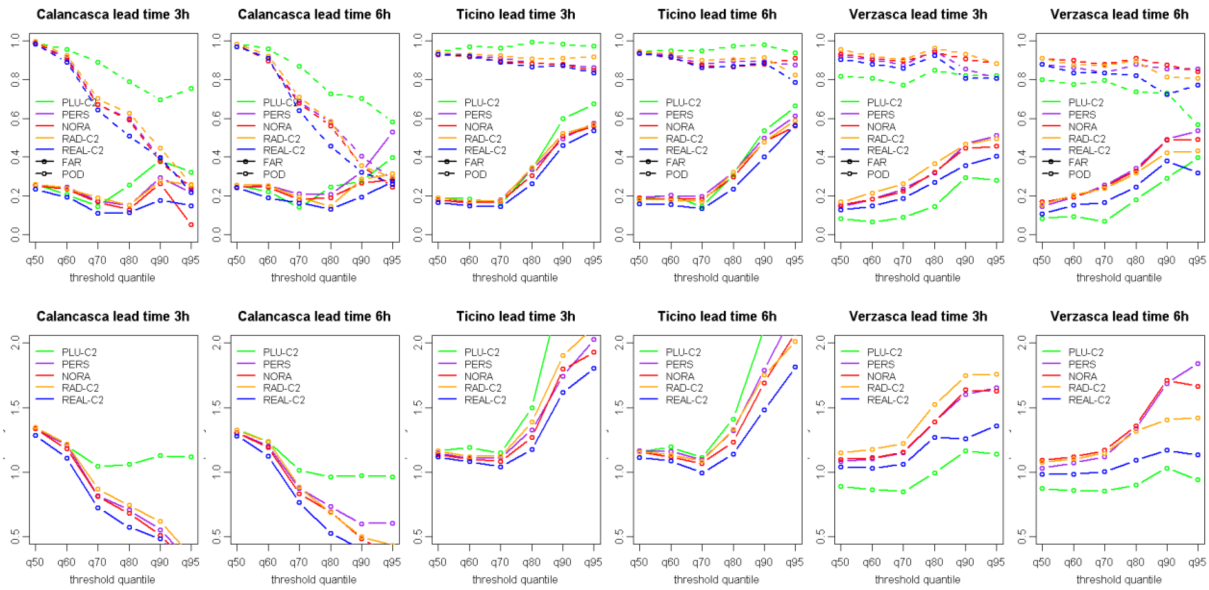

Fig. 5: Probability Of Detection (POD, dashed line), False Alarm Ratio (FAR, solid line) and BIAS (lower panel) for each catchment for q50 to q95 with 3 and 6 hours lead time. Best FAR equals 0, best POD and BIAS equals 1.

6.3 Verzasca

Up to q80 BSS values for REAL-C2 are around 0.6, while for q90 and q95 they are between 0.4 and 0.5. The values generally decrease with increasing lead time. Values for NORA are lower than for REAL, and for q60 and q70, values range between 0.45 and 0.6 with a maximum at 4 and 5 hours lead time. BSS values for q80 are around 0.4 with a maximum at lead time 3 (Fig. 6a). For q90 and q95, BSS values are around 0.2 up to lead times 5 and 6, but then decrease rapidly towards no skill. The persistence (PERS) starts from the same level as with NORA on the shortest lead times (BSS 0.55). However, the skill decreases with increasing threshold (Fig. 6b), and the decrease in BSS over lead time is faster for higher thresholds (Fig. 6a). BSS values for RAD-C2 decrease from 0.5 on q50 to 0.3-0.4 on q70. The BSS for short lead times on q80 are very low, but increase to a maximum of 0.35 for 5 hours lead time. Similar to the persistence, q90 and q95 have no skill on the shortest lead times, however, BSS values show some skill for longer lead times. PLU-C2 reaches BSS values of around 0.6 for q60 to q80, which decrease with lead time (Fig. 6a). The highest BSS value for the shorter lead times (1-4h) was reached with q80 (0.63). For the high thresholds, q90 and q95, BSS values still ranged between 0.4 and 0.5 for lead times of 1 to 3 hours. At short lead times and high thresholds, PLU-C2 keeps up with REAL-C2. If the radar products are compared, scores for NORA are generally below those for REAL, but above those of RAD-C2. For lead times 1 and 2, PERS outperforms RAD-C2 on high thresholds. However, for longer lead times RAD-C2 performs better than PERS. Comparing NORA with PLU-C2, we see that for q50 NORA still scores significantly higher than PLU-C2. This changes for q60 lead time 4, and from q70 onwards PLU-C2 shows better skill than NORA. This difference is most pronounced for short lead times.

All forecast chains show POD values above 0.8 on all thresholds. However, POD and also FAR values for PLU-C2 behave differently in the Verzasca catchment than in the other two catchments. For Verzasca, PLU-C2 shows the lowest FAR and POD values of all forecast chains except q90 and q95, where REAL-C2 is a little bit lower in POD. FAR values generally increase with increasing threshold from about 0.15 to 0.4/0.5. REAL-C2 was the radar product that performed best. With increasing lead time, NORA outperforms RAD-C2 in POD. However, NORA also shows higher FAR values on thresholds higher than 0.6. Furthermore, for longer lead times, PLU-C2 reaches the lowest POD and highest FAR on q95 and not on q90, unlike for shorter lead times. Radar-based forecasting chains show a significant overforecasting bias for q80 to q95. PLU-C2 slightly underforecasts up to q70, and slightly overforecasts for q90 and q95. With increasing lead time, the bias for RAD-C2 becomes smaller for the high thresholds.

6.4 Ticino

REAL reaches BSS values between 0.6 and 0.75 for thresholds between q50 and q70 for all lead times, but then drop significantly, ranging between 0.2 and 0.3 for q90 (Fig. 6b). Furthermore, for q95 REAL-C2 only shows skill for lead times 3 to 6, and event then is below 0.15, i.e. very low. The highest scores for REAL-C2 are reached for q70 (Fig. 6b). For NORA the BSS values between q50 and q70 lay between 0.5 and 0.6 for all lead times. The highest scores are reached for q70. For q80 the values are a bit lower (0.35-0.45) and increase with lead time. For the highest thresholds NORA shows almost no skill. For PERS the highest BSS are reached for q70 at 1 hour lead time (0.6). For both q60 and q70 scores for lead time 1 to 3 are between 0.55-0.6, but then BSS values decrease steadily to 0.4 at lead time 8. PERS show no skill on the highest quantiles (Fig. 6b). BSS values for RAD-C2 for q60 and q70 also range between 0.5 and 0.6, but decrease less with increasing lead time than PERS. For q80 BSS values increase from 0.25 at lead time 1 to 0.33 at lead time 8. Like PERS, RAD-C2 has no skill for q90 and q95 (Fig. 6b). BSS values for PLU-C2 for q50 and q60 decrease with lead time and range from 0.5 to 0.4 and 0.6 to 0.45, respectively. The highest scores are reached for q70 and range between 0.65 and 0.7. For q80 BSS values increase with lead time from 0.32 to 0.42. For the highest threshold quantiles, PLU-C2 shows no skill. For q50, q60 and q80 the skill of PLU-C2 is in the range of NORA and RAD-C2, but for q70 PLU-C2 outperforms all forecast types apart from REAL-C2. In comparison with all the other radar products REAL-C2, shows the most skill. The difference between NORA, PERS and RAD-C2 increases with increasing threshold and longer lead times. NORA performs better than PERS and RAD-C2 on the higher thresholds, but for PERS and RAD-C2 it depends on the lead time. At shorter lead times, PERS scores better and on longer lead times RAD-C2 outperforms PERS.

POD values are high for all thresholds and forecast chains, ranging between 0.75 and 0.99. PLU-C2 shows the highest POD, RAD-C2 the second highest and REAL-C2, NORA and PERS about the same scores. FAR values behave differently and increase rapidly after q70 from about 0.15 to 0.55 and 0.7 respectively. Again PLU shows the highest FAR, REAL-C2 the lowest and the other forecast chains lie in between on about the same level. This matches with the bias obtained for the forecasts in the Ticino catchment. The bias is about 1-1.2 for q50 to q70, and then increases rapidly for all forecasting chains. PLU-C2 is the most biased and REAL-C2 the least over all thresholds. The same behaviour for bias, POD and FAR can be seen when looking at longer lead times, although the POD values for RAD-C2 on q95 are an exception as they are below those for NORA and PERS, and lower than at 3 hours lead time.

The forecast chains are ranked in the same order for Ticino and Calancasca for POD and FAR, but the actual values of POD and FAR behave reversed. For Calancasca, it is POD that worsens rapidly with increasing lead time, whereas for Ticino this is the case for FAR. This is also mirrored in the bias, which mainly shows an underforecasting bias for the Calancasca and an overforecasting bias for the Ticino.

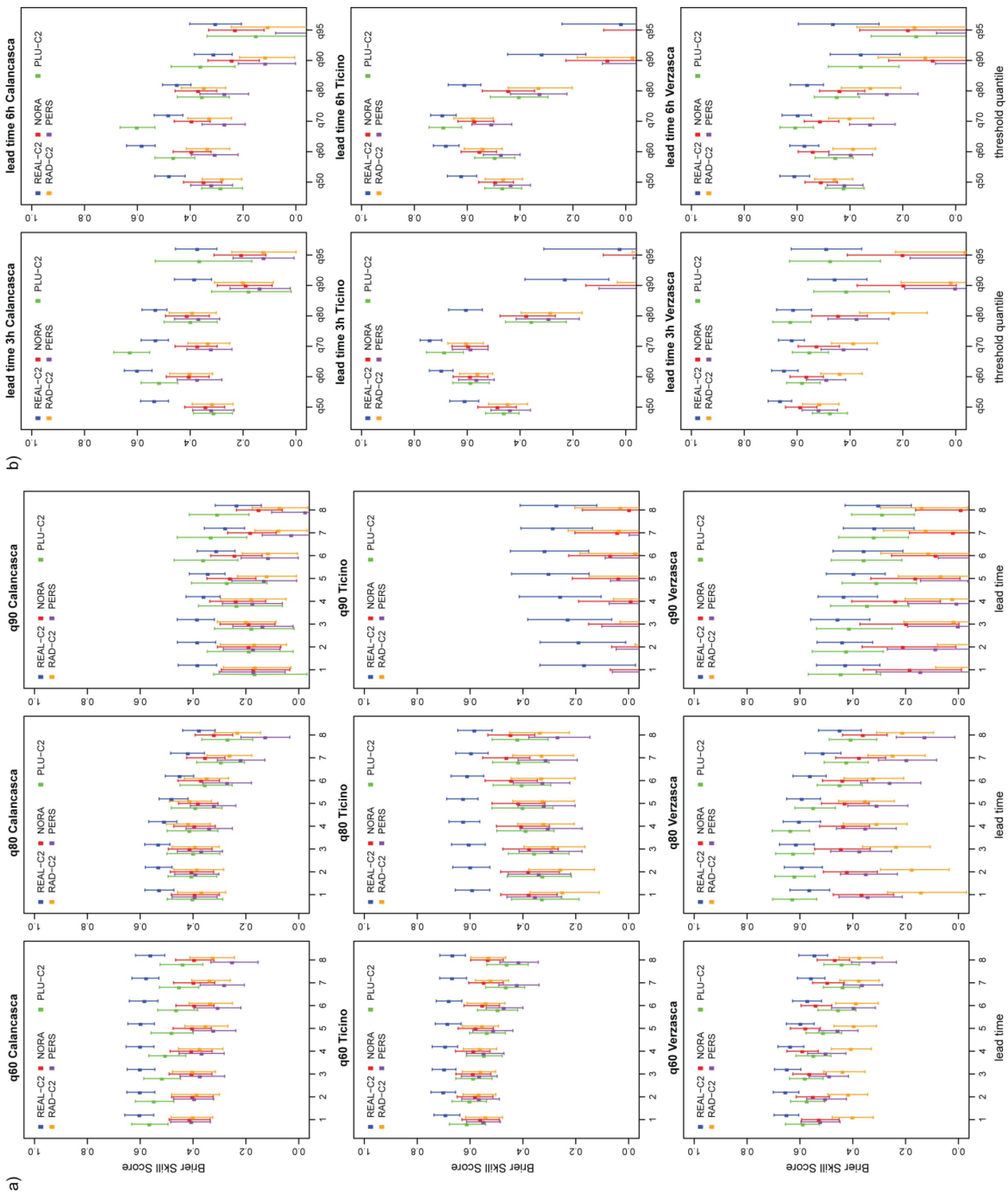

Fig. 6: a) Brier Skill Score (BSS) according to lead time for the threshold quantile q60, q80 and q90. b) Brier Skill Score according to the threshold quantiles q50 to q95 for 3 and 6 hours lead time. Error bars indicate the 95% confidence limits around the estimated BSS value. Positive BSS values indicate an improvement in the forecast over the reference forecast, which in this case is the sample climatology.

6.5 Roc area

The roc areas presented in Table 2 to Table 4 are generally higher than 0.7, which is considered to be the minimum value for a forecast system to be useful for a decision maker (Buizza *et al.*, 1999). For Ticino and Verzasca, they do not drop below 0.9 up to q90. For Calancasca they are a bit lower, especially for NORA and for the high thresholds. For Calancasca and Ticino, the REAL-C2 forecasts have higher roc areas than NORA forecasts on all lead times and thresholds, although this difference decreases with increasing lead time. For the Verzasca catchment, the advantage of REAL-C2 over NORA is only clearly evident on short lead times. Roc areas for REAL-C2 decrease with lead time (except Ticino, q90), but this is not the case for NORA forecasts.

Table 2: ROC area for NORA and REAL-C2 forecasts of the Calancasca catchment, with lead times 3h, 6h and 8h, for the threshold quantiles q60 to q95.

	Calancasca					
	lt3		*lt6*		*lt8*	
	nora	*real*	*nora*	*real*	*nora*	*real*
q60	0.874	0.935	0.877	0.929	0.879	0.924
q70	0.826	0.937	0.853	0.919	0.85	0.899
q80	0.817	0.945	0.839	0.923	0.826	0.899
q90	0.723	0.887	0.764	0.876	0.733	0.830
q95	0.652	0.897	0.736	0.838	0.725	0.825

Table 3: ROC area for NORA and REAL-C2 forecasts of the Ticino catchment, with lead times 3h, 6h and 8h, for the threshold quantiles q60 to q95.

	Ticino					
	lt3		*lt6*		*lt8*	
	nora	*real*	*nora*	*real*	*nora*	*real*
q60	0.911	0.962	0.913	0.959	0.914	0.958
q70	0.920	0.977	0.911	0.960	0.905	0.950
q80	0.902	0.968	0.914	0.967	0.917	0.955
q90	0.905	0.949	0.915	0.951	0.896	0.947
q95	0.934	0.971	0.944	0.968	0.925	0.946

Table 4: ROC area for NORA and REAL-C2 forecasts of the Verzasca catchment, with lead times 3h, 6h and 8h, for the threshold quantiles q60 to q95.

	Verzasca					
	lt3		*lt6*		*lt8*	
	nora	*real*	*nora*	*real*	*nora*	*real*
q60	0.918	0.95	0.915	0.933	0.895	0.918
q70	0.916	0.954	0.923	0.935	0.904	0.911
q80	0.941	0.973	0.939	0.946	0.918	0.910
q90	0.934	0.967	0.911	0.922	0.887	0.888
q95	0.954	0.985	0.942	0.940	0.918	0.927

6.6 Forecast as in operational mode

In our analysis we focused on the performance of the different forecasting chains regarding specific thresholds and lead times. In an operational context the forecasts would be presented as shown in Fig. 7 and Fig. 8. Here the different forecasting chains are shown together and can be visually compared directly. The NORA forecasts are coupled to COSMO-2 forecasts after eight hours, just as with the other forecasting chains at time t0. The examples in Fig. 7 and Fig. 8 show forecasts of an event in the

Calancasca river on June 15 2007 initialised prior to the event and during the event. The NORA forecast prior to the event gives a good estimate of the first peak, which occurred seven hours after the forecast initialisation (t0), but underestimates the main peak, which occurred 21 hours after t0. The REAL-C2 forecast misses this first peak and also underestimates the second peak. For the forecast initialised during the event, however, NORA still underestimates the main peak, but REAL-C2 captures it.

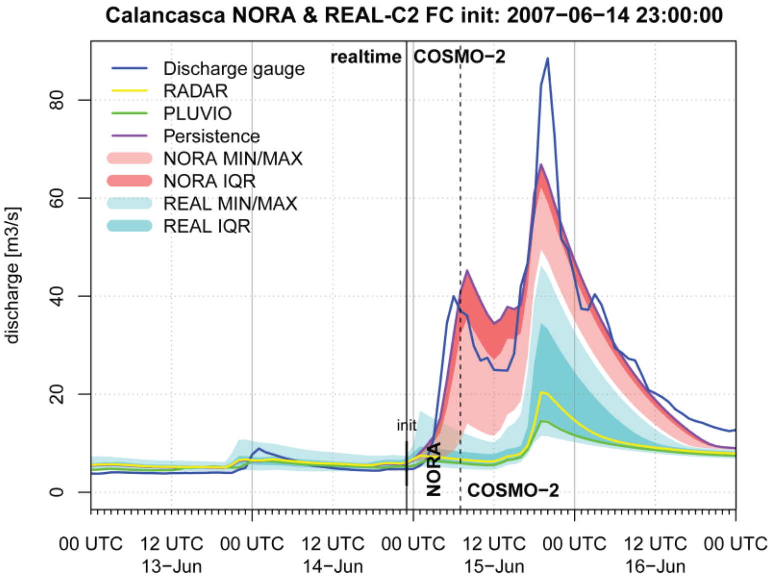

Fig. 7: Forecast simulation for the Calancasca initialized on June 14 2007 at 23:00. At the time of the initialisation of NORA, the nowcasts driven by REAL, RADAR and PLUVIO were coupled to COSMO-2. After eight hours, NORA was also coupled to COSMO-2. The analysis covers the eight hours covered by NORA forecasts, i.e. the time frame between the vertical solid and dashed lines in the graph above.

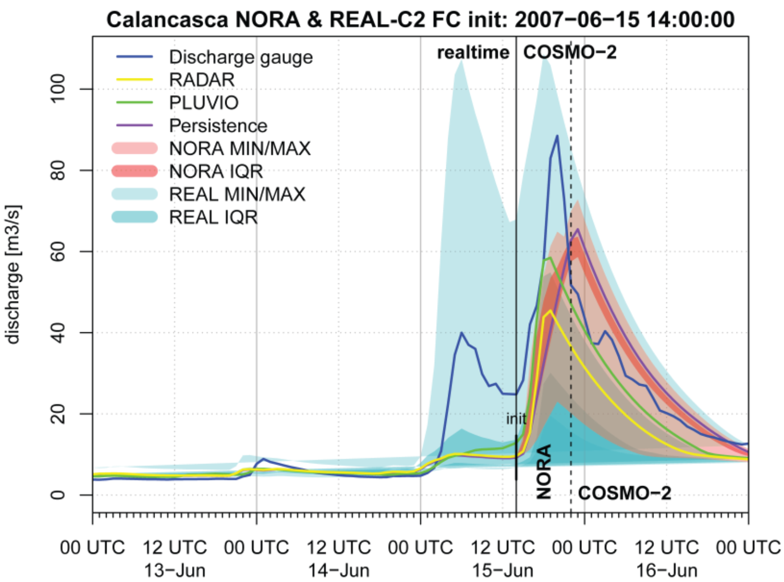

Fig. 8: Forecast simulation for the Calancasca initialized on June 15 at 14:00. For a description, see Fig. 7.

7 DISCUSSION

7.1 Forecast skill

The skills of the different forecast chains, are easier to compare for lower thresholds as the results are clearer and more persistent over the catchments and lead times. Thus general conclusions about the performance of NORA and the other forecasting chains can only be made for quantiles up to q80. For higher quantiles the results vary considerably between the three catchments included in the study.

Panziera et al. (2011) verified NORA for precipitation thresholds of 0.5 and 3 mm per hour, which corresponds to a low threshold that distinguishes between rain and no rain, and a threshold for moderate to heavy rainfall. They integrated their analysis over the whole Lago Maggiore area, but additional investigations for a sub-area showed a similar skill to that for the entire region. They found that NORA performs generally better than Eulerian persistence for the lower threshold, but not for the higher threshold.

Our performance analysis for discharge forecasts shows that NORA also performs better than PERS also for high thresholds, as well as for all catchments on all thresholds over all lead times. Since the events forecast by NORA tend to be persistent by definition, the fact that it also performs better than PERS for short lead times is a valuable finding (Panziera et al. 2011). Moreover, we do not integrate our analysis over the whole Lago Maggiore area, but we analyse the performance of NORA and the other forecasting chains for sub-areas. Thus the variability of precipitation over space and time plays a much greater role in our analysis than in that of Panziera et al. (2011). The good performance of NORA compared to PERS for the discharge forecasts can therefore be explained by the way NORA takes into account the evolution of rainfall (growth and dissipation), whereas PERS keeps the last radar image frozen, and also by the fact that NORA is a probabilistic approach and thus takes into account the uncertainty in the location, time and intensity of precipitation.

Panziera et al. (2011) also compared the performance of NORA with the performance of COSMO-2 for the same rainfall thresholds. For light rainfall (threshold > 0.5 mm/h^{-1}), they found NORA to be better than COSMO-2 for 1-2 hour lead times, and for the higher threshold (3 mm/h^{-1}) NORA outperformed COSMO-2 up to a lead time of 3-4 hours. In the corresponding experiment in our study we compared NORA with RAD-C2, and found that, for thresholds up to q80, NORA generally performs better than RAD-C2 for all lead times, except for Calancasca on lead times 4 to 5 hours. For the Ticino and Calancasca catchments, the 95% interval of the bootstrapped BSS values of NORA and RAD-C2 overlap quite a lot (Fig. 6a), but a Student's T-test on a 5% significance level showed that BSS values of RAD-C2 and NORA are all different apart from Ticino q70 with lead times 4h and 5h, q90 lead time 7h and Calancasca q60 lead time 3h. The advantage of NORA over RAD-C2 up to q80 is, however, clear for the Verzasca basin. For the highest quantiles (q90 and q95), the results differ between the catchments. However, for the Calancasca NORA significantly outperforms RAD-C2 from 4 hours lead time onwards. For Verzasca, on the other hand, NORA performs better than RAD-C2 up to 5 to 7 hours. Forecasts for Ticino basically show no skill on the highest quantile, except NORA on q90, but the BSS values are very small, probably because of the general overforecasting on these high quantiles as the catchment is influenced heavily by water management. It should be noted that the uncertainty of the results for forecast performances is larger the greater the threshold (Fig. 6b) due to the fact that fewer data points lie over the high thresholds. However, in most cases NORA shows higher scores than RAD-C2, which is indicative of the added value in applying a probabilistic approach.

If we combine the advantages of a probabilistic nowcast with COSMO-2, as in the REAL-C2 forecast chain, the result of the comparison looks different. REAL-C2 forecasts perform better than NORA forecasts in all three catchments. This leads to the assumption that given orographic precipitation (or any other repetitive weather situation) plays a major role in flash-flood triggering, and a continuous series of weather radar data is available, NORA can offer a useful method to predict near future discharge. It is especially useful in regions without any ground truth measurements, i.e. rain gauge data, but for regions where REAL can be produced, i.e. in regions where the space-time variance and the auto-covariance of radar errors are known, REAL would be preferred over NORA. REAL also has the advantage that it can be produced continuously and that it is not restricted to orographic precipitation.

The fact that REAL is not only for orographic precipitation is important as an analysis of all exceedances of the q95 threshold from mid June 2007, when the first NORA forecast was made, to December 2010 showed, that a significant part of all threshold exceedances were not covered by a NORA forecast. For Verzasca, 31% of the q95 threshold exceedances lie outside the hours forecast by NORA, for Calancasca 46 %, and for Ticino 42 %. One reason is that the NORA archive contains only situations with orographic precipitation that can be detected by predictors, and excludes local for computational reasons and because spatially and temporally limited events usually do not result in critical situations (Panziera *et al.*, 2011). This may be correct if the whole Lago Maggiore region is considered, but local convective storms can indeed result in extreme discharge events if they remain stationary over a specific catchment.

The advantage of REAL-C2 over NORA is also supported by the ROCA values, which are a measure of the potential skill of the forecasts if the model is correctly calibrated. For the Ticino and Calancasca catchments, the ROCA values for REAL-C2 are always higher than those for NORA. This means that, even if the system has been correctly calibrated with radar data, REAL-C2 would outperform NORA for the current set up of our study. For the Verzasca catchment, REAL-C2 seems to perform better than NORA only on short lead times.

The pluviometer based forecasts PLU-C2 appear to perform remarkably well compared to the other forecasting chains, and in some cases even outperform REAL-C2 (Calancasca q70 and q90 at long lead times, Verzasca q80 at short lead times) (Fig. 6). One reason for this relatively good performance of PLU-C2 could be that the hydrological model was calibrated with rain-gauge data, and this calibration is also the basis for all the model chains based on weather radar data. Furthermore, a bias correction factor is applied to all interpolated rain-gauge data. Thus PLU provides the better initial conditions for the forecasts.

Although the PLU-C2 forecasting chain performs generally relatively well, there are quite some differences between the individual catchments. On the highest quantiles q90 and q95, PLU-C2 has no skill in the Ticino catchment, whereas for the Verzasca and Calancasca catchments it is still skilful. This difference for the Ticino catchment can be explained by looking at the bias of PLU-C2 for the Ticino catchment. The PLU-C2 forecasts for Ticino are very positively biased on the highest quantiles. The other forecasting chains also show a positive bias, but not that extreme, as is also indicated by the high FAR combined with still relatively high POD values for high threshold quantiles. Thus extreme events are overforecasted for the Ticino river, possibly due to the influence of several storage lakes for hydropower production. The precipitation that actually falls in the catchment is not then recorded at the catchment outlet at the estimated time, but is stored in the lakes.

The interpolated precipitation maps are, however, generally good as PLU-C2 performed well in the Verzasca catchment and especially well in the Calancasca catchment, where the PLU-C2 forecasts are mostly unbiased despite the lack of a rain-gauge in the catchment (Fig. 3). However, the good performance of PLU-C2 is also connected to the fact that PREVAH was calibrated using interpolated rain-gauge data.

The forecasts with REAL-C2 are rather negatively biased on the highest threshold quantiles for the Calancasca river, but they are nevertheless still skilful and even outperform the other forecast products on most threshold quantiles and lead times with respect to BSS. This might be explained by the fact that the other radar-based forecasting chains also show a similar negative bias and that the probabilistic approach of REAL-C2 is better than PLU-C2.

The event in the Calancasca river presented in Fig. 7 and Fig. 8 shows that the relatively old COSMO-2 forecast available prior to the event, on June 14 2007 23:00, dampens the performance of REAL-C2. COSMO-2 forecasts only little rain for the first about 15 hours of our forecast, so that the spread of REAL-C2 does not grow much over the first hours. In such situations NORA can help detect critical situations earlier. However, the comparison with the forecast initialised during the ongoing event suggests that the potential of NORA forecasts mainly lies in the early detection of a coming event rather than in forecasting the magnitude of an event, but more individual events need to be analysed, to be able to draw a general conclusion.

7.2 Ensemble spread

Regarding the spread, the different ensembles behave as expected over the eight hours analyzed. NORA forecasts show an increasing spread over lead time. Even though the forcing analogues of the NORA ensemble are very similar, the evolution following this initial time step can be very different, and the possibility of divergence between the individual members increases with each time step as long as there is still precipitation. A NORA forecast also always starts with a single initial state at time t0 (Fig. 4). REAL, on the other hand, is initialized four days prior to the initialization of NORA, by which time it has already built up some spread. Thus the influence of the initial conditions is minor after these four days. At time t0 the REAL nowcast is coupled with the latest available COSMO-2 forecast. This means that the deterministic COSMO-2 is started with 25 different initial states. As the REAL ensemble will have already developed its spread prior to the coupling, the change in spread over the following eight hours is not as big as for the NORA ensemble, which starts from one single initial state. As soon COSMO-2 stops adding more precipitation the ensemble members converge.

The Verzasca is noticeably the only catchment where the spread of the REAL-C2 ensemble decreases with lead time, possibly due to the nature of the events included in the study period. NORA is only produced if the atmospheric conditions favour orographic precipitation, which means in the Lago Maggiore region, that the winds are blowing from the southwest or south (Panziera *et al.*, 2011). Thus storms usually move roughly from southwest to northeast, and therefore arrive and leave the Verzasca basin earlier than the Ticino and Calancasca basin. This also effects the timing of the discharge peaks of the major events within the study period, and the Verzasca river usually peaks at least one hour earlier than the Calancasca river for the events analyzed. The Ticino river also peaks after the Verzasca river, but here the reason is most likely that the Ticino catchment is one order of magnitude bigger than the Verzasca catchment, and thus reacts more slowly.

8 Conclusion

In our study we explored the space between radar-based flash-flood nowcasting and forecasting by evaluating five different flash-flood forecasting systems for three catchments in the Southern Swiss Alps using the hydrological model system PREVAH. Special emphasis was placed on the added value of the purely radar-based NORA forecasting system. NORA is a probabilistic, analogue-based forecast for orographic precipitation, consisting of 12 members, initialized with the initial conditions derived from a four-day nowcast with deterministic radar QPE. The second probabilistic forecasting system evaluated in our study is REAL-C2, where COSMO-2 is initialized with 25 different initial conditions derived from a four-day nowcast with the radar ensemble REAL. Additionally, three deterministic forecasting chains were analysed. One is the persistence of the radar QPE at t0 (PERS), while the other two are COSMO-2 forecasts initialised with initial conditions derived from a four-day deterministic nowcast with radar QPEs (RAD-C2) and interpolated rain-gauge data (PLU-C2).

We analysed the performance of these five flash-flood forecasting systems for all hours between June 2007 and December 2010 for which NORA forecasts were issued, when triggered by orographic precipitation.

We found a clear preference for the probabilistic approach. NORA forecasts outperformed the PERS forecast in all catchments, over all thresholds and for all eight hours lead time. Moreover, NORA generally outperformed RAD-C2 for thresholds up to q80. The best results were, however, mainly achieved with the REAL-C2 forecasting chain, which also showed remarkable skill on the highest threshold. This again demonstrates the advantage of the probabilistic approach. Building an ensemble of 25 different initial conditions with REAL nowcasts leads to better results than NORA, which starts from just a single initial condition.

Further it could also be considered to couple NORA to a REAL nowcast. This would result in an ensemble of 300 members, which would most probably show a very large spread. Thus, to be useful for decision making, some sort of pre-selection of behavioural members would be required. First tests using the Series Distance method (Ehret and Zehe, 2011) to rank REAL members encourage further investigations in this direction. An analysis of this approach would, however, be beyond the scope of the study presented here.

Future investigations may also use NORA forecasts to derive initial conditions for a subsequent initialisation of NWP forecasts, as in REAL-C2. See the example in Fig. 7 and Fig. 8, where COSMO-2 is coupled to NORA after eight hours. The ideal time for coupling NORA to COSMO-2 still needs to be decided. According to the results for the Calancasca and Verzasca catchment, the ideal time for switching from NORA to an NWP forecast would probably be after 4 to 5 hours. This is also in agreement with Panziera et al. (2011) who found that after 4 to 5 hours COSMO-2 precipitation forecasts generally perform better than NORA.

Nowcasts with interpolated rain-gauge data also seem to provide good initial conditions, which could also be used for NORA. However, it would probably be better to calibrate the model with weather radar data, but this would require a long continuous series of weather radar data.

Our study also showed that a well-maintained rain-gauge network is very useful. The forecasts initialised by states derived from nowcasts with interpolated rain-gauge data not only perform very well, but they are also needed to investigate the space-time variance and auto-covariance of radar errors, which is a prerequisite for producing the radar ensemble REAL.

Generally we can conclude that, if the data required to produce REAL are available, REAL-C2 is the preferred forecasting chain because it performs better than NORA and is not restricted to events originating from orographic precipitation. However, for regions where REAL cannot be produced, NORA might be an option to forecast events triggered by orographic rainfall.

ACKNOWLEDGEMENT

We are grateful to MeteoSwiss for permission to use data from their ground-based and weather radar networks, and to the Swiss Federal Office for Environment for providing us with runoff data from their networks. The authors were fully or partly financed through the EU FP7 Project IMPRINTS (Grant agreement no.: 226555 / FP7- 843 ENV-2008-1-226555). Special thanks to Felix Fundel for his support with statistics and programming and to Silvia Dingwall for language editing.

REFERENCES

Addor, N., Jaun, S. and Zappa, M., 2011. An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): skill, case studies and scenarios. Hydrology and Earth System Sciences, 15(7): 2327-2347.

Alfieri, L., Thielen, J. and Pappenberger, F., 2012. Ensemble hydro-meteorological simulation for flash flood early detection in southern Switzerland. Journal of Hydrology, 424: 143-153.

Ament, F., Weusthoff, T. and Arpagaus, M., 2011. Evaluation of MAP D-PHASE heavy precipitation alerts in Switzerland during summer 2007. Atmospheric Research, 100(2-3): 178-189.

Bartholmes, J.C., Thielen, J., Ramos, M.H. and Gentilini, S., 2009. The european flood alert system EFAS – Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts. Hydrology and Earth System Sciences, 13(2): 141-153.

Berenguer, M., Corral, C., Sanchez-Diezma, R. and Sempere-Torres, D., 2005. Hydrological validation of a radar-based nowcasting technique. Journal of Hydrometeorology, 6(4): 532-549.

Buizza, R., Hollingsworth, A., Lalaurette, E. and Ghelli, A., 1999. Probabilistic predictions of precipitation using the ECMWF ensemble prediction system. Weather and Forecasting, 14(2): 168-189.

Efron, B., 1992. Jackknife-after-Bootstrap Standard Errors and Influence Functions. Journal of the Royal Statistical Society Series B-Methodological, 54(1): 83-127.

Ehret, U. and Zehe, E., 2011. Series distance - an intuitive metric to quantify hydrograph similarity in terms of occurrence, amplitude and timing of hydrological events. Hydrology and Earth System Sciences, 15(3): 877-896.

Fundel, F., Walser, A., Liniger, M.A., Frei, C. and Appenzeller, C., 2010. Calibrated Precipitation Forecasts for a Limited Area Ensemble Forecast System Using Reforecasts. Monthly Weather Review, 138(1): 176-189.

Germann, U., Berenguer, M., Sempere-Torres, D. and Zappa, M., 2009. REAL - Ensemble radar precipitation estimation for hydrology in a mountainous region. Quarterly Journal of the Royal Meteorological Society, 135(639): 445-456.

Germann, U., Galli, G., Boscacci, M. and Bolliger, M., 2006. Radar precipitation measurement in a mountainous region. Quarterly Journal of the Royal Meteorological Society, 132(618): 1669-1692.

Germann, U. and Zawadzki, I., 2002. Scale-Dependence of the Predictability of Precipitation from Continental Radar Images. Part I: Description of the Methodology. Monthly Weather Review, 130(12): 2859-2873.

Germann, U. and Zawadzki, I., 2004. Scale Dependence of the Predictability of Precipitation from Continental Radar Images. Part II: Probability Forecasts. Journal of Applied Meteorology, 43(1): 74-89.

Gurtz, J., Zappa, M., Jasper, K., Lang, H., Verbunt, M., Badoux, A. and Vitvar, T., 2003. A comparative study in modelling runoff and its components in two mountainous catchments. Hydrological Processes, 17(2): 297-311.

Jaun, S. and Ahrens, B., 2009. Evaluation of a probabilistic hydrometeorological forecast system. Hydrology and Earth System Sciences, 13(7): 1031-1043.

Liechti, K., Zappa, M., Fundel, F. and Germann, U., 2012. Probabilistic evaluation of ensemble discharge nowcasts in two nested Alpine basins prone to flash floods. Hydrological Processes: n/a-n/a.

Mandapaka, P.V., Germann, U., Panziera, L. and Hering, A., 2012. Can Lagrangian Extrapolation of Radar Fields be used for Precipitation Nowcasting over Complex Alpine Orography? Weather and Forecasting, 27(1): 28-49.

Michelson, D.B., 2004. Systematic correction of precipitation gauge observations using analyzed meteorological variables. Journal of Hydrology, 290(3â€“4): 161-177.

Montani, A., Cesari, D., Marsigli, C. and Paccagnella, T., 2011. Seven years of activity in the field of mesoscale ensemble forecasting by the COSMO-LEPS system: main achievements and open challenges. Tellus Series a-Dynamic Meteorology and Oceanography, 63(3): 605-624.

Morin, E., Jacoby, Y., Navon, S. and Bet-Halachmi, E., 2009. Towards flash-flood prediction in the dry Dead Sea region utilizing radar rainfall information. Advances in Water Resources, 32(7): 1066-1076.

Panziera, L. and Germann, U., 2010. The relation between airflow and orographic precipitation on the southern side of the Alps as revealed by weather radar. Quarterly Journal of the Royal Meteorological Society, 136(646): 222-238.

Panziera, L., Germann, U., Gabella, M. and Mandapaka, P.V., 2011. NORA–Nowcasting of Orographic Rainfall by means of Analogues. Quarterly Journal of the Royal Meteorological Society, 137(661): 2106-2123.

Ranzi, R., Zappa, M. and Bacchi, B., 2007. Hydrological aspects of the Mesoscale Alpine Programme: Findings from field experiments and simulations. Quarterly Journal of the Royal Meteorological Society, 133(625): 867-880.

Rossa, A.M., Laudanna Del Guerra, F., Borga, M., Zanon, F., Settin, T. and Leuenberger, D., 2010. Radar-driven high-resolution hydro-meteorological forecasts of the 26 September 2007 Venice flash flood. Journal of Hydrology, 394(1-2): 230-244.

Rotach, M.W., Ambrosetti, P., Ament, F., Appenzeller, C., Arpagaus, M., Bauer, H.-S., Behrendt, A., Bouttier, F., ois, Buzzi, A., Corazza, M., Davolio, S., Denhard, M., Dorninger, M., Fontannaz, L., Frick, J., Fundel, F., Germann, U., Gorgas, T., Hegg, C., Hering, A., Keil, C., Liniger, M.A., Marsigli, C., McTaggart-Cowan, R., Montaini, A., Mylne, K., Ranzi, R., Richard, E., Rossa, A., Santos, M., oz, D., Sch, r, C., Seity, Y., Staudinger, M., Stoll, M., Volkert, H., Walser, A., Wang, Y., Werhahn, J., Wulfmeyer, V. and Zappa, M., 2009. MAP D-PHASE: Real-Time Demonstration of Weather Forecast Quality in the Alpine Region. Bulletin of the American Meteorological Society, 90(9): 1321-1336.

Sevruk, B., 1996. Adjustment of tipping-bucket precipitation gauge measurements. Atmospheric Research, 42(1â€“4): 237-246.

Szturc, J., Osrodka, K., Jurczyk, A. and Jelonek, L., 2008. Concept of dealing with uncertainty in radar-based data for hydrological purpose. Natural Hazards and Earth System Sciences, 8(2): 267-279.

Tobin, C., Nicotina, L., Parlange, M.B., Berne, A. and Rinaldo, A., 2011. Improved interpolation of meteorological forcings for hydrologic applications in a Swiss Alpine region. Journal of Hydrology, 401(1-2): 77-89.

Velasco-Forero, C.A., Sempere-Torres, D., Cassiraga, E.F. and Jaime GÃ³mez-HernÃ¡ndez, J., 2009. A non-parametric automatic blending methodology to estimate rainfall fields from rain gauge and radar data. Advances in Water Resources, 32(7): 986-1002.

Viviroli, D., Zappa, M., Gurtz, J. and Weingartner, R., 2009. An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools. Environmental Modelling & Software, 24(10): 1209-1222.

Werner, M. and Cranston, M., 2009. Understanding the Value of Radar Rainfall Nowcasts in Flood Forecasting and Warning in Flashy Catchments. Meteorological Applications, 16(1): 41-55.

Weusthoff, T., Ament, F., Arpagaus, M. and Rotach, M.W., 2010. Assessing the Benefits of Convection-Permitting Models by Neighborhood Verification: Examples from MAP D-PHASE. Monthly Weather Review, 138(9): 3418-3433.

Wilks, D.S., 2006. Statistical methods in the atmospheric sciences. Elsevier, Amsterdam, 627 pp.

Wöhling, T., Lennartz, F. and Zappa, M., 2006. Technical Note: Updating procedure for flood forecasting with conceptual HBV-type models. Hydrology and Earth System Sciences, 10(6): 783-788.

Zappa, M., Jaun, S., Germann, U., Walser, A. and Fundel, F., 2011. Superposition of three sources of uncertainties in operational flood forecasting chains. Atmospheric Research, 100(2-3): 246-262.

Zappa, M. and Kan, C., 2007. Extreme heat and runoff extremes in the Swiss Alps. Natural Hazards and Earth System Sciences, 7(3): 375-389.

Zappa, M., Rotach, M.W., Arpagaus, M., Dorninger, M., Hegg, C., Montani, A., Ranzi, R., Ament, F., Germann, U., Grossi, G., Jaun, S., Rossa, A., Vogt, S., Walser, A., Wehrhan, J. and Wunram, C., 2008. MAP D-PHASE: real-time demonstration of hydrological ensemble prediction systems. Atmospheric Science Letters, 9(2): 80-87.

AI

Five perspectives on ranking two severe flash flood events in Switzerland

Katharina Liechti[1], Jens Martin Turowski[1] and Massimiliano Zappa[1]

[1]Swiss Federal Research Institute WSL, Birmensdorf, Switzerland

Abstract:

In June 2007 the atmospheric conditions above Switzerland were characterized by flat pressure gradients that triggered several stable and spatially limited convective rainfall events. Here, we analyse two such events, which are particularly well documented. The studied sites are the Langeten catchment in Canton Berne and the Einsiedeln region with the Alp and Minster sub-catchments in Canton Schwyz in Switzerland.

The events have been analyzed from five different perspectives: a) Estimation of the economic value of the damages on the basis of the Swiss flood and landslide damage database; b) Analysis of the daily, hourly and 10-minutes precipitation with extreme-value statistics (GEV); c) Analysis of the hourly and 10 minutes discharge rates from nested basins with GEV; d) GEV analysis of sediment transport in the Erlenbach catchment using geophone data. e) Hind casting the events with the hydrological model PREVAH using operationally available quantitative precipitation estimates from both pluviometers and weather radar.

For all time series analysed with GEV the two events rank amongst the most severe in their respective regions, and were responsible for the highest damage costs. The estimated return periods for the event in the Langeten catchment were > 100 y for the 10 minute runoff and 70 y for daily precipitation. For the Alp catchment the return periods were estimated to 70 y for 10 minute runoff, 35-40 y for 10 minute precipitation and 40 y for sediment transport. The results illustrate known problems with extreme value analysis, such as the occurrence of threshold processes in runoff generation.

Hydrological prediction of flash flood events can be improved if weather radar observations are used to constrain the precipitation events. However, the quality that can ultimately be achieved depends on the extent of the thunderstorm, and on the spatial and temporal resolution of the radar measurements. A possible route for improvements may lie in the combination of radar and ground-based observations.

Flash floods are still insufficiently understood and documented. To optimize flood hazard mitigation and aid decision makers, we call for a Europe-wide flash flood atlas and data base.

INTRODUCTION

Flash floods are characterized by a strong response to intense rainfall and rapidly rising and falling hydrographs, resulting in limited opportunity for warnings to be prepared and issued (Perry, 2000; USGS 2008; Collier, 2007). Even though flash floods triggered by intense storms cause huge economic damage and loss of life every year in the whole of Europe, they are still a scarcely documented and poorly understood natural phenomenon (Gaume et al., 2009; Rusjan et al., 2009). In Switzerland, the data compiled in the Swiss flood and landslide damage database (Hilker et al., 2009) show that thunderstorms are responsible for about 25% of the total damage caused in natural hazard events, amounting to approximately 1850 million Euros since 1972 (taking inflation into account).

Since extreme value analysis is frequently used to rank flash flood events and assess their severity, high quality long-term data series are required for rigorous inspection. However, often such events occur in ungauged catchments where no data is recorded, or their limited spatial extent is incorrectly captured by conventional monitoring such as meteorological or river observation networks (Rusjan et al., 2009). In addition, when data is collected about a specific event, it often remains unpublished and frequently the material disappears into the archives of local authorities, companies or research units, where it is difficult to find and access (Gaume et al., 2009).

Few studies report high-quality observations of individual flash floods. Griffiths et al. (2009) analysed precipitation data collected with both pluviometers and radar for flash floods causing extensive landsliding and debris flows in southeastern Arizona in July 2006. Rusjan et al. (2009) described an extreme rainfall event and the subsequent flash floods in Western Slovenia in September 2007. They estimated storm return period applying the Gumbel distribution to rainfall data and confirmed that post-flood investigation should focus on discharges and hydrological response rather than on rainfall. A more general approach is to use the generalized extreme value distribution (GEV) which unifies the three original families of extreme value distribution into a single equation (Coles, 2001). In practice, this distribution has been used to model a wide variety of natural extremes such as floods, rainfall, wind speeds, and wave height (Martins and Stedinger, 2000).

As the main triggers of flash floods are the spatio-temporal distribution, the duration and the intensity of rainfall. The major challenge for forecasters is to accurately provide this information. Often rainfall data from rain gauges captures the spatial distribution and intensity of a rainfall event insufficiently (Griffiths et al., 2009; Rusjan et al., 2009). Delrieu et al. (2009) suggested that weather radar offers a unique means for characterizing the temporal and spatial variability of rainfall with the resolution required for a large variety of hydrological problems. Especially for risk management in catchments with fast response like mountainous or urban catchments, radar rainfall estimates are a promising tool (Borga et al., 2002; Krajewski et al., 2002; Morin et al., 2009). Borga et al. (2007) performed hydrometeorological analysis of a flash flood in the Eastern Italian Alps in August 2003. To achieve this they utilised precipitation estimated from high resolution weather radar and from rain gauge networks, flood response observations derived from stream gauge data and post-event surveys. They recognized and analyzed two main controls on extreme flash flood response: steadiness of convective bands and dry antecedent soil moisture conditions. Weather radar provides a convenient means to monitor the first of these. However, as the radar network is improved there seems to be a trend of reducing the density of traditional rain gauge networks due to the high effort and costs required for installation and maintenance (Delrieu et al., 2009).

In 2007 several flash flood events occurred in northern Switzerland. Two of these were extensively monitored and high-quality data series are available for precipitation (both pluviometer and weather radar), runoff and sediment transport. The high-quality data in addition to field investigations of the affected catchments provide the rare opportunity to analyse such spatially limited, hydrometeorological extreme events from five different perspectives:

a) Estimation and ranking of the economic value of the damage, on the basis of the Swiss flood and landslide damage database of the Swiss Federal Research Institute WSL (hereafter referred to as WSL; Hilker et al., 2009);
b) Analysis of the daily, hourly and 10-minutes precipitation with extreme-value statistics using the general extreme value distribution (GEV);
c) Analysis of the hourly and 10 minutes discharge rates from nested basins with GEV;
d) Analysis of the accumulation of sediments at the Erlenbach basin (a small headwater basin within the Alp). Geophones records available since 1986 allow GEV analysis of sediment transport.
e) Hind casting the events with the hydrological model PREVAH (Viviroli et al., 2009a) using operationally available data from both pluviometers and weather radar (Germann et al., 2006).

The objectives of this study are, (1) to obtain a better understanding of magnitude and impact of the events, (2) to estimate and compare the return period of the different processes involved (rainfall, runoff, sediment transport), (3) to assess the presumed advantage of data source diversity for the post-event analysis, and (4) to analyze the possibilities of forecasting flash floods on the basis of operationally available data, in particular weather radar quantitative precipitation estimates (QPE, Germann et al., 2006). The variety of available observations and data types for our events makes a broad and comprehensive analysis possible.

TWO FLASH FLOOD EVENTS IN SWITZERLAND

General situation and meteorology

In Switzerland the first half of June 2007 was dominated by flat pressure gradients leading to several severe and local convective rainfall events. Due to low winds the convective cells moved particularly slowly, generating extreme thunderstorms at local sites (MeteoSwiss, 2007).

The first of the analyzed events occurred on June 20th in the Einsiedeln region (Figure 1). Here, after a hot summer day, a cluster of thunderstorms formed over the central Swiss Pre-Alps. The western part of this cluster moved slowly towards Alptal and Lake Sihl, later towards the upper part of Lake Zürich. In the Einsiedeln region precipitation rates of 72.5 mm per hour were recorded by MeteoSwiss. Several tributary streams of the Alp and the Minster were breached due to the heavy rainfall. Total damage of ~ 39 million Euros was caused by flooding and severe sediment transport.

The second event analysed here occurred on June 8th in the Langeten catchment (upstream of the gauge in Huttwil; Figure 2). In this region, the humid and unstable air masses are forced to ascend by the relief of the Napf that pertrudes out to the Swiss midland (Napf region), making the area prone to thunderstorms. Even so, the event from June 8th can be considered as extraordinary. According to weather radar estimates of MeteoSwiss around 100 mm rain fell within a few hours, an amount which normally accumulates over a whole month. This led to severe flooding causing three fatalities and damages of ~ 40 million Euros in three municipalities.

Catchments and available data

The Alp catchment and Einsiedeln region

The Einsiedeln region is situated in the central Swiss Pre-Alps in Canton Schwyz. It includes the Alp and Minster catchments, each draining into the river Sihl (Figure 1). Four municipalities are located in the region: Einsiedeln and Alpthal in the Alp catchment, and Oberiberg and Unteriberg in the Minster catchment. Five pluviometers operated by MeteoSwiss measuring daily precipitation are located within this region (see Figure 1 for their locations). The east and west side of the Alp catchment is bordered by hill ranges of about 1500 m height. To the South, the large and small Mythen mark the highest points in the catchment at 1800 and 1900 m a.s.l. Towards the North the valley is open with the town of Einsiedeln near the confluence with the river Sihl (Burch, 1994). At the gauging station in Einsiedeln, the Alp catchment drains an area of 46.4 km^2, with an average elevation of 1155 m a.s.l. and an average slope of 40.8 m/km. The main stream has a total length of 16.4 km. Geologically, the Alp catchment can be divided into three sections. The Mythen chalk cliffs delimit the catchment in the South. Different groups of Flysch can be found in the centre of the valley, and to the north molasses dominate (BAFU, 2008a).

In three tributaries of the Alp, the Erlenbach, the Lümpenenbach and the Vogelbach, discharge is gauged and pluviometers provide rainfall measurements at 10 minute intervals (Hegg et al., 2006). In the case of the Erlenbach catchment, the pluviometer is a part of a climate station also providing other relevant parameters (air temperature, water vapour pressure, global radiation, wind speed, sunshine duration and precipitation) for hydrologic modelling. The runoff in the Lümpenenbach was not unusual in 2007 and is not discussed here. In the Vogelbach, a landslide caused a power cut at the gauging station and debris blocked the measurement channel. The recorded data is thus unreliable and cannot be used to estimate the return period of the event. However, from geomorphic evidence it is clear that it was one of the most severe events since the start of observations in the late seventies.

The Erlenbach has been intensively monitored by WSL since 1978 (Burch, 1994; Hegg et al., 2006). A discharge time series is available, spanning 28 years at 10-minute resolution. In 1982, a sediment retention basin was constructed, which has been surveyed regularly for the deposited volumes. In 1986 indirect bedload sensors were installed in the channel bed upstream of the basin. These were replaced and upgraded in 1999. From November 2002 onwards, after an initial testing and calibration phase, reliable information was recorded again (Rickenmann and McArdell, 2007; Turowski et al., 2009). The time series used for the extreme value analysis of bedload transport consist of the largest recorded event in 1984 and the yearly maxima from 1986 to 1999 and from 2002 to 2008.

The Langeten catchment

The Langeten catchment is located in the Swiss central plateau, upstream of the town of Huttwil in Canton Berne. It encompasses an area of 59.9 km^2, including the tributaries Rotbach and Wyssachen (Figure 2). It comprises the three municipalities Huttwil, Eriswil and Wyssachen. The average elevation of the catchment is 766 m a.s.l., with the highest point at 1119 m a.s.l. in the southeast. The length of the main stream Langeten up to the Huttwil gauging station is 10.2 km, with an average slope of 36.6 m/km. Geologically, the catchment can be divided into two parts. The southern part lies in the Napf conglomerate whereas the northern part, with the gauging station in Huttwil, lies in upper fresh water molasses (BAFU, 2008b). For the Langeten 33 years of 10 minute runoff measurements are available for statistical analysis of the 2007 event at the Huttwil gauging station. Daily precipitation has been monitored since 1972 at two rain gauges within the Langeten catchment at Huttwil and at Eriswil. Nearby stations which deliver hourly or 10 minute precipitation were not directly affected by the storm and therefore cannot be used for analysis.

For both catchments rain gauge and climatological data required for runoff modelling were provided by the Swiss national automatic measurement network (ANETZ), which operates on 10 minute intervals. Additionally weather radar QPE was provided by MeteoSwiss in the framework of the MAP D-PHASE project (Zappa et al., 2008). Further information on the rain-gauge network is provided in section 3.2.

The Swiss flood and landslide damage database

Since 1972 data on damage caused by naturally triggered floods, debris flows, landslides and (since 2002) rockfall events in Switzerland have been collected in the Swiss flood and landslide damage database by the WSL (Hilker et al., 2009). The main source for the database are ~ 3000 Swiss newspapers and magazines, which are scanned for information about damage caused by the processes listed above. Ideally, these provide some basic information on the meteorology of the events, the type of process involved and an estimate of the total damage cost. If the latter is not included it is estimated by the WSL. The database includes damage data for the municipalities in the studied catchments. The studied events are compared to and ranked with respect to records of historical events from the database.

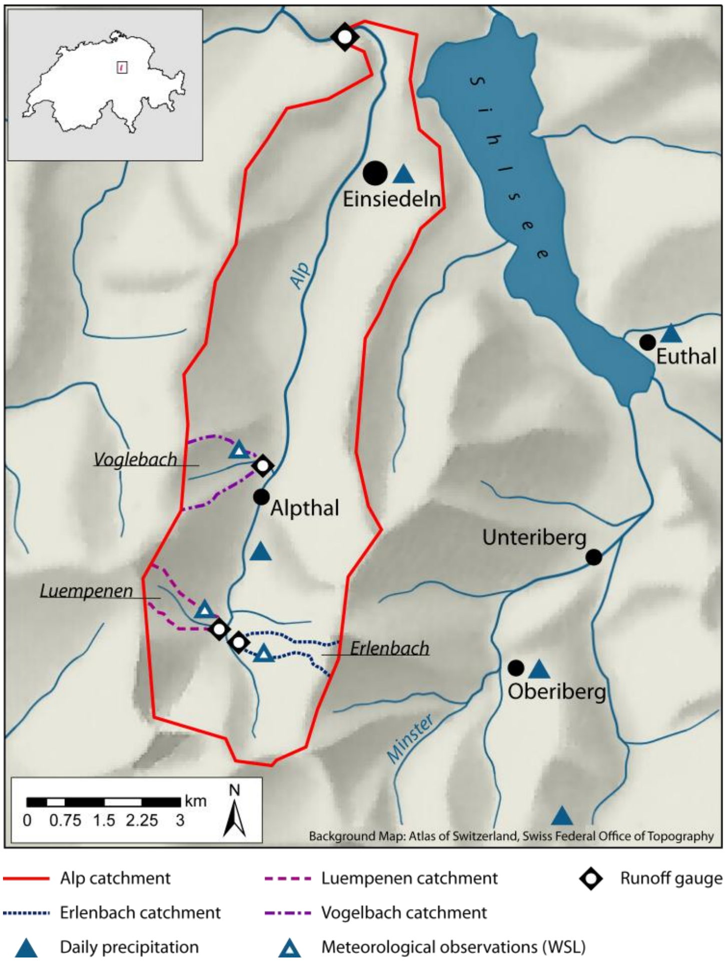

Figure 1: Alp catchment with its three small tributary catchments and hydrometeorological observing networks used.

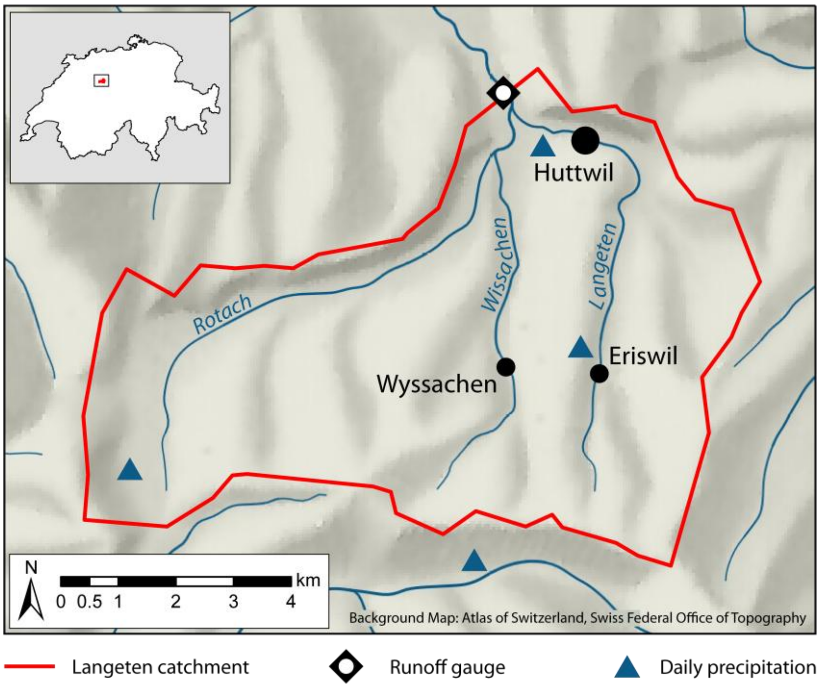

Figure 2: Situation of the Langeten catchment including location of the observing discharge and precipitation stations used.

METHODS

Extreme value statistics

For the return period analysis we chose the generalized extreme value distribution (GEV) introduced by Jenkinson (1955). It incorporates three types of commonly used extreme value distributions in a single function, the Gumbel distribution (Type I), the Fréchet distribution (Type II) and the Weibull distribution (Type III) (Martins and Stedinger, 2000). These three types show different behaviours. The Weibull distribution shows an upper boundary, the Fréchet distribution a lower boundary, whereas the Gumbel distribution has neither. The standardisation of the three types simplifies the extreme value analysis substantially. No subjective decision is needed as the data itself defines the matching type of extreme value distribution through the shape parameter (Leadbetter et al., 1983).
Three parameters define the GEV: location μ, scale σ and shape ξ. It reduces to one of the three distribution types for different ranges of the shape parameter: to the Gumbel distribution with $\xi=0$, the Fréchet distribution with $\xi > 0$ and the Weibull distribution with $\xi < 0$. The cumulative distribution function of the GEV is written as:

$$G(x) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \qquad \text{for } \xi \neq 0 \qquad (1)$$

$$G(x) = \exp\left\{-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\} \qquad \text{for } \xi = 0 \qquad (2)$$

where $1+\xi(x-\mu)/\sigma>0$ and $-\infty<\mu<\infty$, $\sigma>0$ and $-\infty< \xi <\infty$ (Coles, 2001).

The parameters of the GEV were estimated by the maximum likelihood method (Aldrich, 1997). As the GEV does not satisfy the regularity conditions required for the usual asymptotic properties associated with the maximum likelihood estimator to be valid, Smith (1985) gave the following guidelines:

- $\xi < 0.5$, the maximum likelihood estimation is valid, it has the usual asymptotic properties.
- $1 > \xi > 0.5$, the maximum likelihood estimation has a result, but the standard asymptotic properties are not fulfilled.
- $\xi > 1$, maximum likelihood estimation is unlikely.

For extreme values $\xi \geq 0.5$ is rare. Therefore, the theoretical restriction of the maximum likelihood method is normally no obstacle in practice (Coles, 2001). For the analysis of series of yearly maxima from natural processes shape parameters above 0.5 indicate problems with the data or different processes involved.

For analysis the block maxima approach was chosen and yearly maxima were taken from the data series. Prior to the analysis the series of yearly maxima were tested against homogeneity, stationarity and independence. As no yearly maxima lie close to the end or beginning of a year the samples are considered to be independent. Homogeneity and stationarity were tested by trend analysis using the Mann-Kendall-test (Kendall, 1970) and the Run-test (Wald and Wolfowitz, 1940). The tests showed no trend for all samples at a significance level of 0.05.

Hydrological modelling

The runoff hind casts were realized with the semi-distributed hydrological model PREVAH (Precipitation Runoff Evapotranspiration HRU related model) (Gurtz et al., 2003; Zappa et al., 2003). The model works in hourly time steps and on a 500 x 500-m grid. This grid is aggregated by the delineation of hydrological response units (HRUs) taking into consideration information on topography, land use and soils depth (Gurtz et al., 1999). PREVAH is forced by hourly hydrometeorological data. Different sources and methods have been applied to obtain spatially distributed rainfall input. Operationally accessible rain gauge data from the ANETZ network were spatially interpolated for the target regions by the inverse distance weighting method (Viviroli et al., 2009a). Additionally, for the target area “Alp”, rainfall data from a local network run by WSL consisting of three pluviometers (Hegg et al., 2006) were used as additional data points for precipitation interpolation (see also Figure 1 and section 2.2.). Weather radar QPE (Germann et al., 2006) were downscaled to force the hydrological model (Germann et al., 2009; Zappa et al., 2009). An application of PREVAH with a similar configuration (pre-processing and calibration) is described in more detail by Zappa and Kan (2007).

For the Langeten the time series analyzed with PREVAH span from 1981 to 2007, according to the availability of hourly meteorological data. The simulations for the Alp cover the whole period of available observed discharge data from the years 1991-2007. The calibration of PREVAH was completed following the procedure presented in Viviroli et al. (2009b). For the model calibration, data from the ANETZ network were used (air temperature, precipitation, water vapour pressure, global radiation, wind speed and sunshine duration). With the inverse distance weighting technique these data were interpolated to build meteorological surfaces. For air temperature, water vapour pressure, global radiation and wind speed an elevation-dependent adjustment factor was applied (Jaun and Ahrens, 2009; Zappa and Kan, 2007) . Additionally daily observations from a dense network of pluviometers were available, these are the only source of information on precipitation amounts for the Langeten catchment measured directly in the catchment. The Langeten catchment was calibrated with eleven different calibration periods, the Alp catchment with six, each consisting of five consecutive years. The use of different calibration periods results in a small ensemble of parameter sets and allows a quantification of uncertainty (Pappenberger and Beven, 2006). The year prior to the calibration period was used as an initialization year. The year 2007 was not included in any of the periods defined for model calibration.

Uncertainty of the obtained calibrated parameters was evaluated with a Monte Carlo analysis (Viviroli et al., 2009b). Skill scores for 1000 randomly generated parameter sets were compared to the scores achieved by the parameter sets resulting from the model calibrations (for more detail see Appendix). Full information of the setup, physics and calibration procedure for PREVAH can be found in a comprehensive review of the PREVAH modelling system by Viviroli et al. (2009a and 2009b).

RESULTS

Alp catchment and Einsiedeln region

Precipitation Alp

The daily rainfall amounts measured at the gauges in the Einsiedeln region on June 20th 2007 represent no significant deviation from the norm and occur regularly, with a return period of one to two years. Only the daily rainfall accumulation of Oberiberg shows a slightly higher return period of about seven years (Table 1).

At the rain gauge in the Erlenbach catchment we found return periods of 35 years for both hourly and 10 minute precipitation intensities. In the Lümpenenbach, on the western side of the valley just opposite the Erlenbach, precipitation rates were much lower with a return period of about three years. In the Vogelbach subcatchment 3 km downstream from the Erlenbach on the western side of the valley hourly and 10 min precipitation were estimated to have a return period of 25 and 40 years respectively (Table 2).

Table 1: Daily precipitation sum (NS) for the storm on June 20th 2007. The estimated return period (T) and the 100 year daily cumulative precipitation are indicated for every gauging station (NS100).

Cumulative daily precipitation	***Start of rec.***	***NS100***	***NS*** $_{20.6.2007}$	***T*** $_{20.6.2007}$
Einsiedeln	1900	125	37.9	1
Alpthal	1973	160	72.5	2
Oberiberg	1959	175	95.8	7
Euthal	1973	145	48.7	1
Hoch Ybrig	1973	150	41.1	1

Table 2: Parameters for the fitted GEVs and 50-year return period for maximum precipitation intensity (NS50 [mm]). 10 minutes and 1 hour cumulative values are analyzed separately. The return period T for the records on June 20th 2007 for the three high resolution rain gauges within the Alp are estimated.

Peak intensity	***Series***	***Location*** μ	***Scale*** σ	***Shape*** ξ	***NS50***	***NS*** $_{20.6.2007}$	$T_{NS\ 20.6.2007}$
Erlenbach 1h-sum	1982-2007	21.3	7.0	0.081	54	50	35
Erlenbach 10min-sum	1982-2007	11.8	3.5	-0.006	25	24	35
Lümpenen 1h-sum	1987-2007	24.6	7.7	-0.241	44	29	2
Lümpenen 10min-sum	1987-2007	12.5	3.6	-0.311	21	15.3	3
Vogelbach 1h-sum	1986-2007	20.9	6.6	0.304	70	55.8	25
Vogelbach 10min-sum	1986-2007	11.2	2.8	0.185	27	25.9	40

Table 3: Parameters for the fitted GEVs, 50-year return period for peak runoff (HQ50) and values of June 20th 2007 with corresponding return period T for the Alp.

Yearly peak runoff	*Series*	*Location* μ	*Scale* σ	*Shape* ξ	*HQ50*	$Q_{20.6.2007}$	$T_{Q\ 20.6.2007}$
Hourly average	1991-2007	56.7	17.6	0.009	126	124	43
Eff. Hourly average	1991-2007	58.3	18.5	0.128	152	161	70
10-minutes average	1991-2007	63.6	18.8	0.270	193	213	70
10-minutes average	1991-2006	65.4	17.9	-0.217	112	213	Inf

Runoff Alp and Erlenbach

The 10 minute discharge peak observed at the Alp gauge was 213 m^3/s. The comparison of the hourly and 10 minute runoff data from the Alp for that event indicates that the period of highest runoff is divided onto two hours. Thus, instead of the measured hourly runoff means, the six largest consecutive 10 minute runoff records in each year was used for analysis of hourly runoff data. Return periods for these “effective” hourly runoff means and the 10 minute runoff means were estimated to ~ 70 years (Table 3). It is observed that omitting the extreme value 2007 while estimating the GEV parameters changes the type of distribution function from Fréchet to Weibull. Weibull distributions are bounded at the top, i.e. they are able to describe data up to a maximal value. The extreme value 2007 lies above this upper limit making an estimation of its return period impossible (Table 3).

Even though the data series for the Erlenbach catchment (Figure 1) is complete and the runoff measurements of the 2007 event are reliable, it is very difficult to estimate its return period. The runoff value of the 2007 event (12.9 m^3/s) is clearly an outlier in the 28 years of measurements (Figure 3). With a value of 0.99 the shape parameter of the GEV is very high. This implies that other runoff generation processes than usual have been activated and were dominant during the flash flood event (see 3.1). Similar outliers were recorded during the floods in 1984 (10.2 m^3/s) and 1995 (10.1 m^3/s). The significance of such threshold processes is discussed later.

Erlenbach sediment transport

The 1984 event in the Erlenbach transported far the most sediment with 2230 m^3 deposited in the retention basin. This corresponds to an estimated return period of 55 years. The event 2007 with 1630 m^3 was the second biggest event with a return period of 40 years. While in third, the 1995 event deposited 385 m^3 of sediment and gave an estimated return period of ~ 7 years (Figure 4).

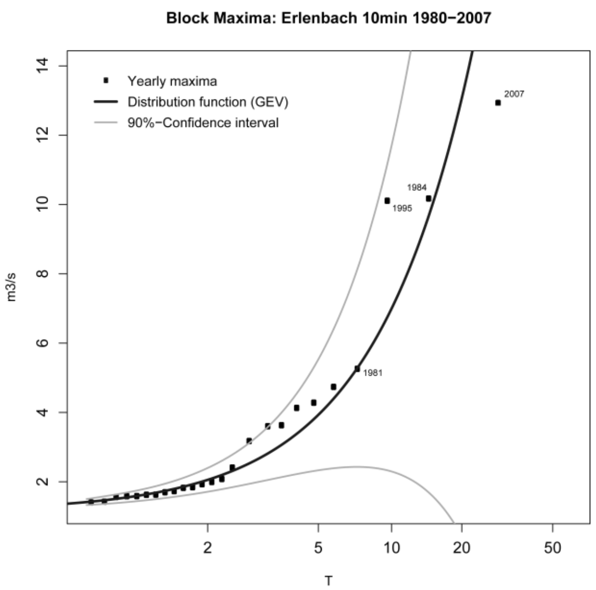

Figure 3: Gumbel-Diagram of the Erlenbach 10 minute peak runoff yearly maxima.

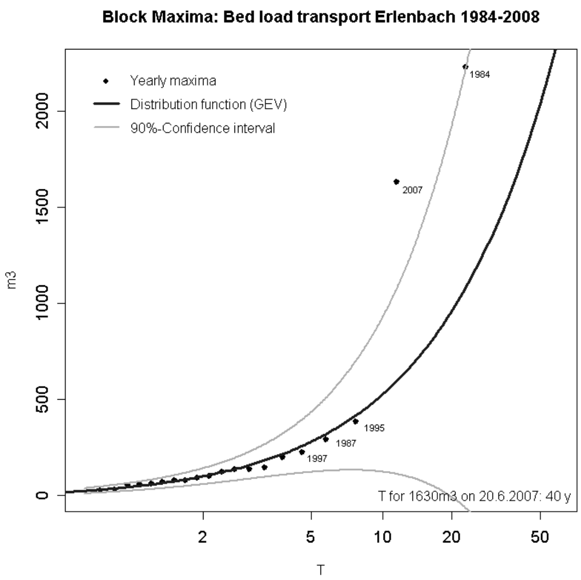

Figure 4: Gumbel-Diagram of the Erlenbach bed load transport yearly maxima, bed load transport in m^3 per event. No data for years 1985 and 2000.

Alp Modelling

The hind casts on the basis of the interpolated ANETZ rain data result in large underestimation of peak runoff (Figure 5). This is not that astonishing since the average station density is 1 station per 570 km^2 and thus too coarse to appropriately capture convective precipitation events. Remarkably, the hind casts using the rainfall data of the WSL network show the same underestimation, even though the network is with one station per 15 km^2 very dense. With peak flows of 23 m^3/s to 33 m^3/s these simulations predict only 18-26 % of the observed hourly mean and only about 10-15 % of the observed 10 minute peak flow.

In contrast the hind casts based on weather radar data (1 km^3 spatial resolution, 10 minute temporal resolution) achieved better results. The ensemble captures the observed hourly mean (124 m^3/s) well. Regarding the observed 10 minute peak flow (213 m^3/s) there is still an underestimation but not as grave as with point data. Five out of six model runs, with peak runoffs between 87 m^3/s and 104 m^3/s, lie in the range of about 70-85 % of the observed hourly mean (124 m^3/s) and 40-50 % of the observed 10 minute peak flow (213 m^3/s). One model run shows a peak runoff of 137 m^3/s, which is 110 % of the observed hourly and 65 % of the observed 10 minute runoff maxima.

Einsiedeln region damage

The Swiss flood and landslide damage database from WSL was analysed for the Einsiedeln region, including the Alp and Minster catchments. Substantial damage costs were recorded in the region in 27 out of 36 years (75% of the time), marking the region as vulnerable to natural hazards (Figure 6). The 2007 event in the Einsiedeln region caused costs of damage of ~ 39 million Euros, contributing 8 % to the total cost of damage 2007 in Switzerland. This is 15 times greater than the average yearly damage costs in the Einsiedeln region.

In two of the four municipalities (Einsiedeln, Unteriberg) the heavy damage makes the 2007 event the most severe in 36 years. In the other two municipalities the costs are of a similar magnitude as the long term average. The three heaviest events regarding the damage (1984, 1990, and 2007) were all caused by thunderstorms triggering landslides and flash-floods. These storms differed in spatial extent and in duration.

The event on July 29th 1990 was spatially the most concentrated of the three. In Oberiberg (Minster catchment) a daily precipitation sum of 173 mm was observed, whereas at the gauging stations 6 km away in the neighbouring Alp catchment only 30 mm was recorded. In Euthal 6 km downstream the valley from Oberiberg daily precipitation summed to only 62.5 mm. On July 25th 1984 several weather systems leading to storms hit the Alp catchment yielding twice the daily precipitation sum as on June 20th 2007. This resulted in damages exceeding all damages experienced from past and subsequent events for the Alpthal municipality.

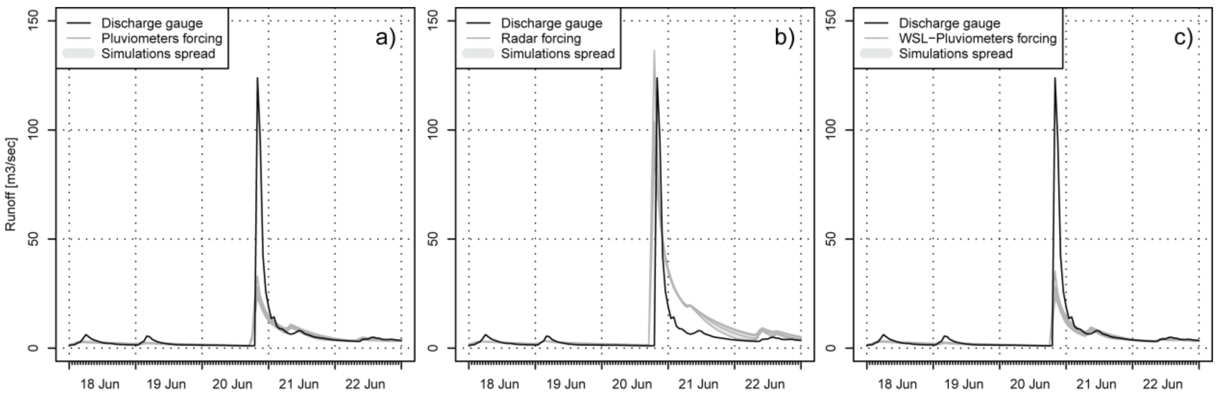

Figure 5: Alp basin,18th to 22nd June 2007: hourly discharge simulations with PREVAH forced by a) pluviometers (ANETZ network), b) by weather radar QPE and by c) a local raingauge network (WSL). The observed hourly runoff is plotted in black. The light-grey area represents the spread obtained by applying 6 different sets of calibrated model parameters. The single model runs are plot as dark-grey lines.

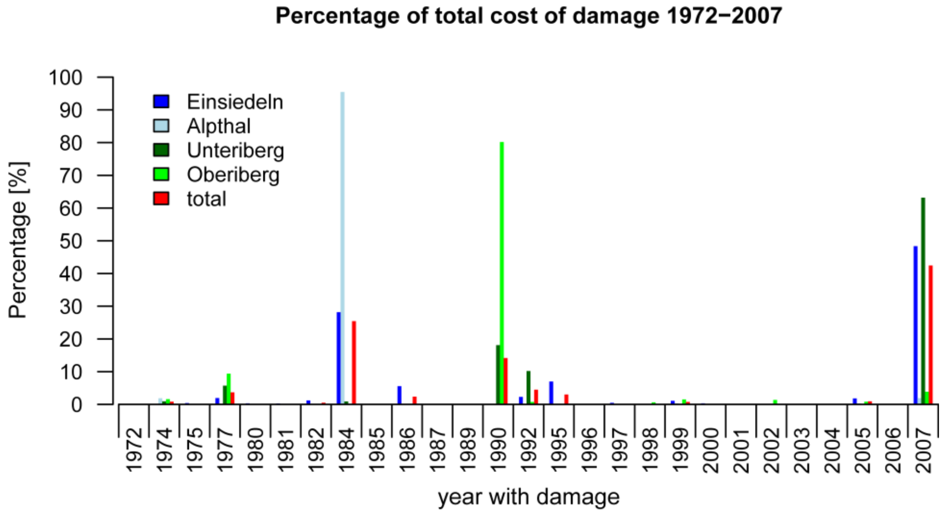

Figure 6: Yearly damage costs in the Einsiedeln Region as percentage of the corresponding total damage cost for the years 1972 to 2007. No bar: no damage respectively damage below 0.1 % of the total damage costs. Only years with recorded damage are shown.

Langeten catchment

Langeten Precipitation

A comparison of the yearly maxima of the precipitation in Huttwil and Eriswil shows that the dates of the yearly maxima do not match, even though Huttwil is located only 4 km downstream of Eriswil in the same valley. In only one case do the two records correspond, this is seen for a long precipitation event in November 1972. In general, the yearly runoff maxima correspond to larger precipitation events, recorded either in Huttwil or in Eriswil.

With a total of 4.8 mm precipitation measured in Huttwil on June 8th 2007 an estimation of return period is not reasonable, as such low values occur many times within a single year. At the rain gauge in Eriswil however, the highest daily sum ever observed (90.5 mm) was measured on this day.

Using yearly maxima from the gauge at Eriswil, the return periods of this event was estimated by fitting the GEV to the time series 1972-2006 and 1972-2007. As the 2007 maxima is much higher than the previous maxima (the second highest daily precipitation observed at the station was 74.7 mm), it has a big influence on the shape of the GEV and therefore on estimating the return period. We estimated the return period of the 2007 precipitation to be 70 years when it is included in the time series and 450 years otherwise (Table 4).

Langeten Runoff

The measured runoff of the Langeten on June 8th 2007 is extraordinary and can be considered to be an outlier in the series of yearly maxima. To see the impact of this event on return period estimation the GEV was fitted to the series with and without this value (Table 5). The shape parameters are much higher after including the value 2007 leading to higher estimated probability for high runoff values. Table 5 shows that a 100 year flood is estimated to be about one third higher after including the 2007 event into the time series. The difference is even more striking considering the estimated return period of the event of June 8th 2007. Including the event into the time series, its return period is estimated at 115-130 years; omitting it while fitting the GEV the return period is estimated strikingly higher at 460-580 years (Table 5).

Table 4: Parameters for the fitted GEVs and 100-year return period for daily cumulative precipitation (NS100 [mm]) for two gauges within the Langeten basin. 10 minutes and 1 hour cumulative values are analyzed separately. The records for June 8th 2007 are declared with the corresponding estimated return period T.

Daily rainfall	***Series***	***Location*** μ	***Scale*** σ	***Shape*** ξ	***NS100***	$NS_{8.6.2007}$	$T_{NS\ 8.6.2007}$
Eriswil 24h	1972-2007	47.9	10.0	0.006	95	90.5	70
Eriswil 24h	1972-2006	48.0	9.8	-0.121	85	90.5	450
Huttwil 24h	1972-2007	45.2	11.2	0.117	115	4.8	nn

Table 5: Parameters for the fitted GEVs, 100-year return period for peak runoff (HQ100) and values of June 8th 2007 with corresponding return period T for the Langeten.

Yearly peak runoff	***Series***	***Location*** μ	***Scale*** σ	***Shape*** ξ	***HQ100***	$Q_{8.6.2007}$	$T_{Q\ 8.6.2007}$
Hourly runoff	1974-2007	13.4	5.8	0.258	64	68	120
Eff. Hourly runoff	1974-2007	13.7	5.9	0.244	64	68	130
10-minute runoff	1974-2007	14.6	6.6	0.265	74	77	115
Hourly runoff	1974-2006	13.3	5.3	0.147	48	68	500
Eff. Hourly runoff	1974-2006	13.6	5.5	0.134	48	68	580
10-minute runoff	1974-2006	14.5	6.1	0.154	55	77	460

Langeten Modelling

Similar to the results from the modelling of the Alp, the hind casts resulting from interpolated ANETZ precipitation underestimate the runoff peak (Figure 7). The simulated runoff peaks at ~20 m^3/s, which is only 30% of the observed hourly peak runoff (68 m^3/s). This underestimation can also be seen for the two smaller peaks on June 10th and 11th, which are barely realized in the hind casts. In contrast, using radar data as input, these little peaks are rather overestimated, but the peak runoff on June 8th is underestimated as before. During normal runoff periods the hind casts simulate the observations well (Figure 7).

Huttwil region damage

Ranking the event of June 8th 2007 in terms of damage, the Swiss flood and landslide damage database was analyzed for the three municipalities Huttwil, Eriswil and Wyssachen, which comprise the Huttwil region within the Langeten catchment. It is remarkable that for the municipalities in the Langeten catchment damage costs were incurred for only 10 out of 36 years despite the high susceptibility to thunderstorms in this region. The damage of June 8th 2007 caused costs of ~ 40 millions Euros in the whole catchment (Figure 8), which is 8.5 % of the total 2007 damage costs in Switzerland, ranking it the highest costing event for the region since the start of data acquisition. The individual municipalities within the catchment were affected with different intensity. Damage costs of the event of June 8th 2007 go beyond the scale of what has been observed before for the whole of the region, and for the municipalities Eriswil and Wyssachen in particular (Figure 8). For these two municipalities the damage costs yield more than thirty times the yearly average of the period 1972-2007. Huttwil experienced high damage too, but these only rank second behind the hazardous impacts of the event of August 30th 1975, which was also characterized by high intensity rainfall and affected a larger region than the 2007 event.

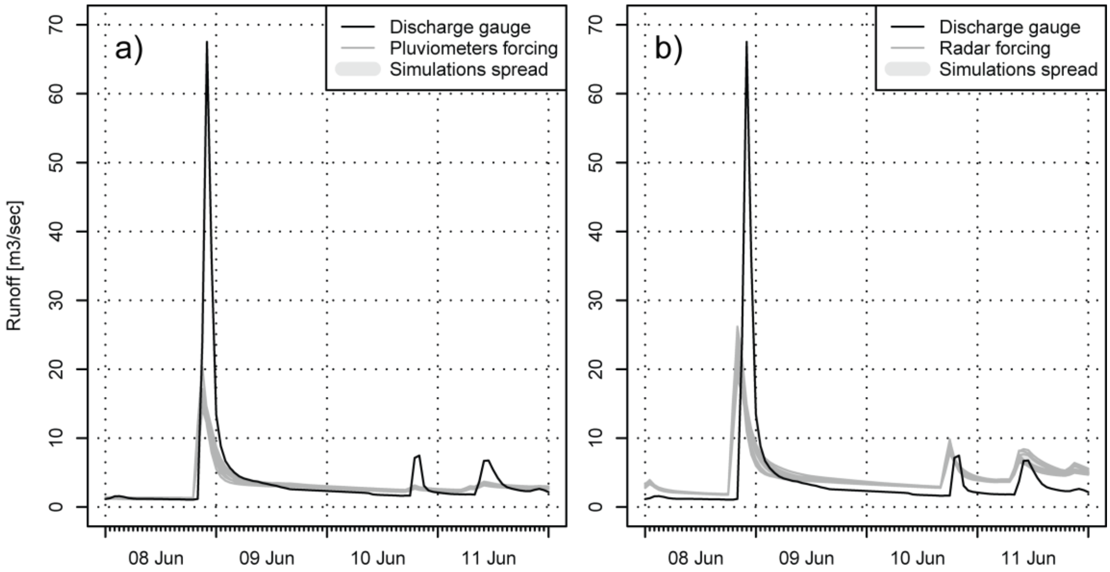

Figure 7: Langeten basin, 8th to 11th June 2007: hourly discharge simulations with PREVAH forced by a) pluviometers (ANETZ network) and by b) weather radar QPE. The observed hourly runoff is plot in black. The light-grey area represents the spread obtained by applying 11 different sets of calibrated model parameters. The single model runs are plot as dark-grey lines.

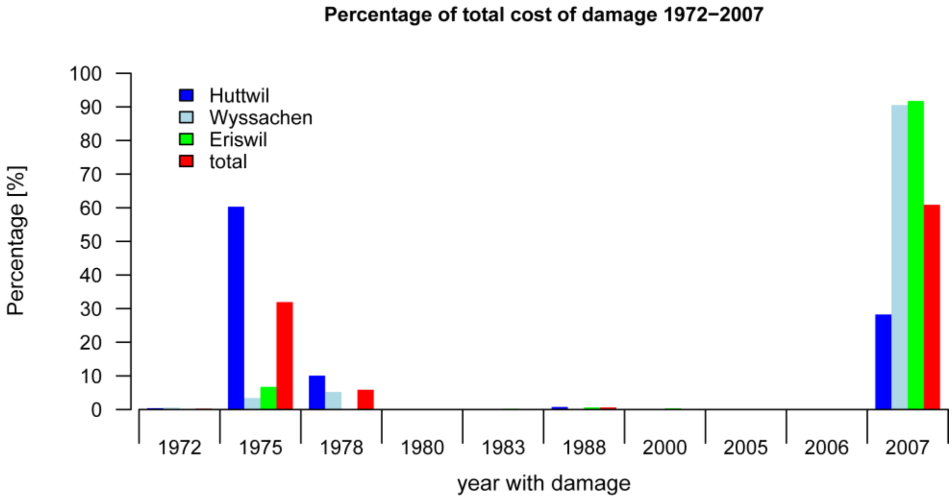

Figure 8: Yearly damage costs in the Huttwil Region as percentage of the corresponding total damage cost for the years 1972 to 2007. No bar: no damage respectively damage below 0.1 % of the total damage costs. Only years with recorded damage are shown.

Discussion

The results of the extreme value analysis and the hydrologic modelling generate many questions about the accuracy, reliability and predictive value of extreme value statistics and hydrological modelling given the operationally available information. Below we first discuss the potential errors and pitfalls of the statistical analysis. We then outline how runoff extremes and extreme values of bed load transport are related to each other. Last, we address the value of weather radar measurements for the modelling of catchment response.

Extreme value statistics

Applying extreme value statistics to our comparably good data series we can identify well known problems concerning accuracy and applicability (*cf*. Merz and Blöschl, 2008). The temporal resolution of the data series has to be of adequate accuracy for the process to be studied. The different return periods estimated for the daily precipitation in Alpthal, and the temporally better resolved data from the WSL rain gauges, illustrate and confirm that coarse temporal resolution of measurements entails the underestimation of return periods in case of flash flood events. Thus, a high temporal resolution is of paramount importance for the analysis of flash floods.

The Langeten and Erlenbach runoff analysis are good examples that illustrate the limits of applicability of the GEV and of extreme value analysis in general. The extremes measured in June 2007 turn the whole extreme value statistics upside down. In conventional analysis, such outliers are often removed from the series and replaced by the second highest value of the same year (DVWK, 1999). But as the observed extremes are not a measurement error but plausible values it would be unjustifiable to remove them from the series, a decision that is also supported by the guidelines of DVWK (1999).

In case of the Langeten we can argue from two sides. First, the estimated return period clearly exceeds three times the length of the time series, which is assumed to be a reasonable limit for the useful estimation of return periods (DVWK, 1999). Second, even though the estimated shape parameter of the GEV lies below 0.5, which implies that the values are statistically sound according to Smith (1985), the three highest yearly maxima do not fall on the trend defined by the rest of the data points. This deviation is most probably the result of the onset of threshold processes in runoff generation within the basin. These processes are most likely determined by interaction of soil hydraulic properties, soil moisture and rainfall distribution and intensity of the event (Figure 9) leading to overland flow (Zehe and Sivapalan, 2009).

In case of the Erlenbach the shape parameter of the fitted GEV exceeds 0.5. This implies that the fitted distribution function is not reliable enough to estimate the return period for the 2007 event at the Erlenbach, clearly indicating that threshold processes are involved in runoff generation. Based on observed soil moisture conditions in the catchment (IHW, 1998), it can be assumed that the infiltration and retention capacities of all areas in the catchment is exceeded at a certain spatial concentration and intensity of the rainfall event, leading to quick surface runoff in the whole catchment area, including both hortonian and saturated overland flow (e.g., Scherrer and Naef., 2003).

In principle a separate analysis with the outliers caused by threshold processes could be performed. However, for such an analysis there are not enough observations exceeding the thresholds.

In conclusion, it can be stated that for the Erlenbach the event of June 20th 2007 was the largest since the start of monitoring, and that the estimated return periods for the Langeten have to be taken with caution.

These findings are particularly relevant for local administrations and practitioners that rely on return period estimations when planning the realization of flood mitigation measures. Infrastructure with high importance to society is recommended to be protected to 100% from floods with return period of 20 years (BWG, 2001). It is obvious that an accurate estimation of the dimension of a 20-year flood is required for achieving realistic cost estimations. On the one hand, an overestimation of the dimension of the 20-year flood may lead to construction costs beyond what can be afforded by the local authorities. On the other, if threshold processes in runoff generation play a role in the catchment, the dimension of the 20-year flood could be underestimated, resulting in insufficient protection for the infrastructure in question. Consequently, high damage costs would result following a flash-flood event.

The implied existence of threshold processes in runoff generation in the Erlenbach catchment is mirrored in the observed bedload transport in the Erlenbach. For both runoff and bedload transport the extreme values in 1984 and 2007 are clearly outliers. 1984 is the biggest event regarding bedload transport, whereas for runoff the event 2007 is ranked first. The 1995 event is ranked in third place for both processes. For runoff this event marks an outlier, whereas for bedload transport it does not. This shows that extreme runoff does not necessarily imply extreme bedload transport. Bedload transport is commonly thought to be a monotonically rising function of discharge, but also depends on the amount of available sediment (e.g., Rickenmann, 2001; Warburton, 1992). Sediment availability depends on the recent history of flooding (which empties storage by transporting material out of the catchment) and of geomorphic events such as landsliding (which refill storage). Furthermore, a major event can change stream bed morphology causing longer-lasting changes in bedload transport processes. Large discharges exert large forces which can destabilise the bed for a considerable length of time. Consequently, sediment transport rates are elevated after heavy flooding (e.g., Gintz et al., 1996). In the Erlenbach, Turowski et al. (2009) observed an increase in bedload transport rates after both the 1995 and 2007 events (no data is available for the 1984 event). This increase was attributed to a destabilisation of the channel bed, leading both to increased sediment availability in the stream and decreased energy losses due to friction, and thus increased sediment transport capacity. The situation becomes worse if return times for bedload transport are to be estimated purely from precipitation time series. Adding this level applies additional uncertainties because it does not take into account the non-linearities in the hydrologic response. Thus, the multiplicity of possible combination of conditions makes it very challenging to conclude from a driving process to the behaviour of a dependent process. Return time estimation of bedload transport events in ungauged or partially gauged catchments, which are for example necessary for the dimensioning of retentions basins, may thus be unreliable.

Value of radar measurements for operational hydrology

The accurate characterization of rainfall fields, both in spatial extent and in temporal evolution, is currently one of the biggest challenges in hydrometeorology and even more difficult in mountainous areas such as Switzerland (Germann et al., 2006). The spatial variability of rainfall and the interaction between weather systems and topography makes even basic estimations very difficult. The thunderstorm of June 8th 2007 affecting the Langeten catchment is an impressive example for this. Over a distance of only 4 km a difference in precipitation of almost 90 mm was observed. Often thunderstorms, which are the main triggers of flood events in the Langeten catchment, are only recorded by one of the two rain gauges, either in Huttwil or in Eriswil. This implies that flooding is caused by rainfall cells which may only cover a few percent of the total catchment area.

Likewise, thunderstorm precipitation did not stand out in the series of daily rainfall accumulations both in the Einsiedeln region and in the Langeten catchment, unsurprisingly showing that extreme precipitation often is of short duration. Consequently, very different results are achieved when hourly or 10-minute precipitation data is used for analysis rather than daily data.

These examples show that even a dense network of monitoring stations has limited potential for the analysis of short, intense rainfall events if it lacks the adequate temporal resolution. In fact, by comparison to most other regions, with three stations the WSL network in the Alp catchment is both very dense and has an excellent temporal resolution. In addition, daily precipitation sums from a rain gauge in the village of Alpthal is available. Such high-quality networks are rare and the fact that the storm directly hit this area can be seen as pure chance. But as comparisons of hind casts based on interpolated ANETZ and WSL rain gauge data show, even a network with high spatial and temporal resolution is no guarantee for representative values over a whole catchment area and the predictive quality of the measurements is limited. Weather radar operates both at high temporal and spatial resolution and can easily cover large areas. As can be seen from the hind casts for the Alp event, simulations based on radar data can achieve much better results than simulations based on ground observations (Figure 5). Therefore, radar rainfall estimates are a promising tool in hydrological modelling of events for which precipitation is the main driver.

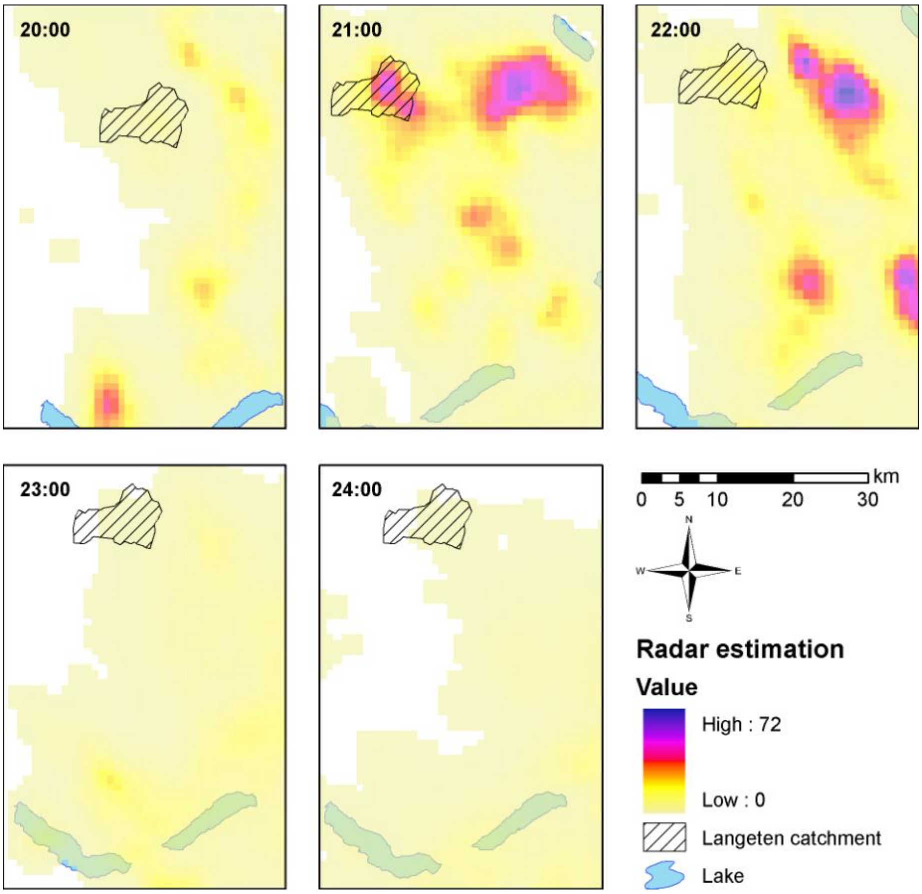

Figure 9: Sequence of hourly weather radar quantitative precipitation estimates [mm/h] for the event on June 8th 2007 in the Langeten catchment. Radar data courtesy from MeteoSwiss.

Nevertheless, it would be wrong to say that radar data is the solution of the problem. The hind cast for the event on June 8th 2007 in the Langeten catchment serves as an example to illustrate that radar data hold serious uncertainties, too, and are not in every case capable to achieve better results. The success of radar data for the prediction of extreme events depends very much on the spatial extension of the convective cell and the location of the respective catchment (considering error sources like radar shading). In case of the thunderstorm in the Langeten catchment the extension of the cell was just too small to be accurately captured by the weather radar. Evidence for that is given in Figure 9, where a sequence of radar precipitation estimates over five subsequent hours is shown. The devastating storm emerges and disappears within a very short time and is concentrated over a very limited area.

A possible improvement in precipitation estimation could lie in the combination of point measurements from rain gauges and data from weather radar (Schiemann et al., 2011). To obtain absolute precipitation values, the weather radar needs to be calibrated and corrected. Despite significant improvements in the last decade the current weather radar QPE are still less accurate than ground-based measurements (Germann et al., 2006). Point data from rain gauges can be used as ground truth to adjust weather radar data (Fuentes et al., 2008). In this way, the advantages of both systems can be combined. The high accuracy of the rain gauges together with the high spatial resolution and coverage of the weather radar can contribute to better precipitation estimates. Another approach would be the use of uncertainty estimates of the quantitative radar precipitation measurements. Such an approach for generating weather radar rainfall ensembles for use in hydrology was proposed by Germann et al. (2009).

Another clear limitation of the adopted hydrological modelling approach is the use of an hourly time scale. The prediction of flash floods like the ones occurred in our target areas would surely benefit from a model being able to process meteorological information at a higher time resolution than one hour. Furthermore, conceptual models like PREVAH generally simplify the runoff generation process by defining a chain of storage elements that are generating surface runoff, interflow and baseflow (Zappa and Gurtz, 2003). A runoff generation module relying on the representation of dominant runoff processes (Scherrer and Naef., 2003; Schmocker-Fackel et al., 2007) would probably be better suited for coping with threshold processed in runoff generation.

CONCLUSIONS

We have described in detail two flash flood events triggered by intense thunderstorms in the summer of 2007 in Switzerland. The affected areas were equipped with high-quality meteorological and hydrological observatories, which allowed an analysis and statistical ranking of the events from five different perspectives: (i) economic damage caused; (ii) precipitation statistics; (iii) runoff statistics; (iv) bedload transport; and (v) hydrological hind-casting.

In both regions the analysed events ranked amongst the most severe in all five categories since the start of records. However, estimations of flood return periods from precipitation data are often inaccurate. Flood hazard mitigation measures, for example sediment retention basins, are often designed based on a 50 or 100 year return flood. Although the extreme value curves of precipitation, discharge and transported sediment volumes compare well in general in the Erlenbach, extreme discharge does not necessarily lead to extreme sediment transport. Our analysis shows that estimation of return periods for sediment transport from precipitation or discharge time series needs to be taken with great care. Unfortunately, long-term data series on transported sediment volumes are rare and a general conclusion cannot currently be made.

Despite the excellent quality of observations, hydrological hind-casts of the flood events show that ground-based data is not sufficient to make accurate operational predictions, for warning purposes. Radar data can improve the quality of predictions, however, thunderstorms are often limited in spatial and temporal extent and are often insufficiently resolved by the instruments. Thus, radar remains a promising tool for operational hydrology, but it will not be the solution to all problems. Radar measurements can be calibrated and validated with ground-based rainfall observations, underlining the importance of maintaining a high-quality and high-density rain-gauge monitoring network.

With damage costs totaling 470 million Euro in Switzerland, the year 2007 can be considered a high damage year, when compared to the mean and median of 235 million and 60 million Euros per year from 1972 to 2006, respectively. The extraordinary high damage costs of 2007 significantly increase the long term means. As a consequence, former hazard events seem less accentuated in the series, even if they caused heavy damage, like the events in the Langeten catchment on August 30th 1975 and July 11th 1978. Taking into account the limited spatial extent of the 2007 events in the Huttwil and Einsiedeln region the percentages of 8.5% and 8% of the total damage costs 2007 in Switzerland become even more impressive.

These observations demonstrate that documentation and analysis of flash flood events and archiving of results in central databases is an important issue as it is the basis of politically and economically viable strategies for flood-disaster reduction. Comprehensive, standardised and georeferenced information on floods is essential for decision-making, monitoring and assessment (Barredo, 2007). In addition, centralized national and international databases storing meteo-hydrological, hydraulic and socio-economic data related to past flash flood events would greatly assist the understanding of flash flood magnitude and occurrence (Creutin and Borga, 2003). Gaume et al. (2009) compiled an inventory of 550 flash floods in seven European regions. A European wide view on major flood disasters between 1950 and 2005 with some focus on flash floods was presented by Barredo (2007). These are valuable steps towards building comprehensive data bases and a European atlas of flash flood events. Moreover, the importance of the phenomenon in terms of damage costs makes the need for further efforts into this direction clear.

ACKNOWLEDGMENTS

MeteoSwiss is acknowledged for providing data from their ground-based and weather radar networks. We thank the Swiss Federal Office for Environment for providing us with runoff data from their networks. Ch. Frei from MeteoSwiss provided great support in implementing the GEV statistics. A. Badoux and N. Hilker from WSL provided data of the Swiss flood and landslide damage database. Many thanks go to N. Addor and J. Glover being great help as graph designer and lector. We thank the reviewer of a previous form of this manuscript for a thorough and constructive review. K. Liechti is financed through the EU FP7 Project IMPRINTS (Grant agreement no.: 226555 / FP7- 843 ENV-2008-1-226555). J.M. Turowski was supported by NSF Grant 200021_124634/1. M. Zappa profited from funding of the State Secretariat for Education and Research SER (**C05.0105**) in the Framework of the COST 731 Action.

APPENDIX

To evaluate the goodness of the parameter sets resulting from automatic model calibration of Vivroli et al. (2009b) a Monte Carlo analysis was performed to generate 1000 random parameter sets which then were evaluated for each calibration period. The efficiency achieved by the parameter sets is shown by the Nash-Sutcliffe efficiency (NSE) score (Nash and Sutcliffe, 1970) and the sum of weighted absolute errors (WAE) (Lamb, 1999). The NSE measures the improvement of the model relative to the mean of the observation (Viviroli et al., 2009b) and is defined as follows:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(O_i - S_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2}, \quad NSE \in [-\infty, 1] \qquad \text{(A.1)}$$

where O_i is the observed runoff at time i, $\overline{O}$ the mean of observed runoff over the entire simulation period of length n, S the simulated runoff at time i. The NSE tends towards 1 when S tends towards O_i, for NSE of 0 no improvement is made compared to the mean of the observation and NSE < 0 means that the simulation performs poorer than the mean of the observation (Wilcox et al., 1990).

The WAE is more sensitive to the errors in the simulation of high flows. It is defined as follows:

$$WAE = \sum_{i=1}^{n}\left(O_i^a \left|O_i - S_i\right|\right), \quad WAE \in [o, \infty[\qquad \text{(A.2)}$$

Lamb (1999) proposed an a of 1.5 which increases the sensitivity to peak flows without becoming too insensitive to recession and low-flow periods. The lower the WAE value the better the peak flow periods are simulated.

The Monte Carlo runs were ranked by a total acceptability score described in Viviroli et al. (2009b). This score is focussed on obtaining robust parameter sets for both peak-discharges and low-flows. The simulations with the calibrated parameter sets achieve an average NSE of 0.76 for the Alp and 0.66 for the Langeten. The average WAE is 1900 for the Alp and 50 for the Langeten. The scores achieved using the parameter sets resulting from the calibrations lie, with only one exception, in the same range as those of the top 50 Monte Carlo runs (Figures A1 and A2). This result show that our model calibration is well representative for simulating both peak-discharge and low-flows. Nevertheless, if we would have focussed on parameter sets being best-suited for peak-discharge, then, especially for the Alp basin, there are numerous Monte Carlo realizations with more favourable NSE/WAE than the ones we obtain from the standard calibration procedure adopted following Viviroli et al. (2009b).

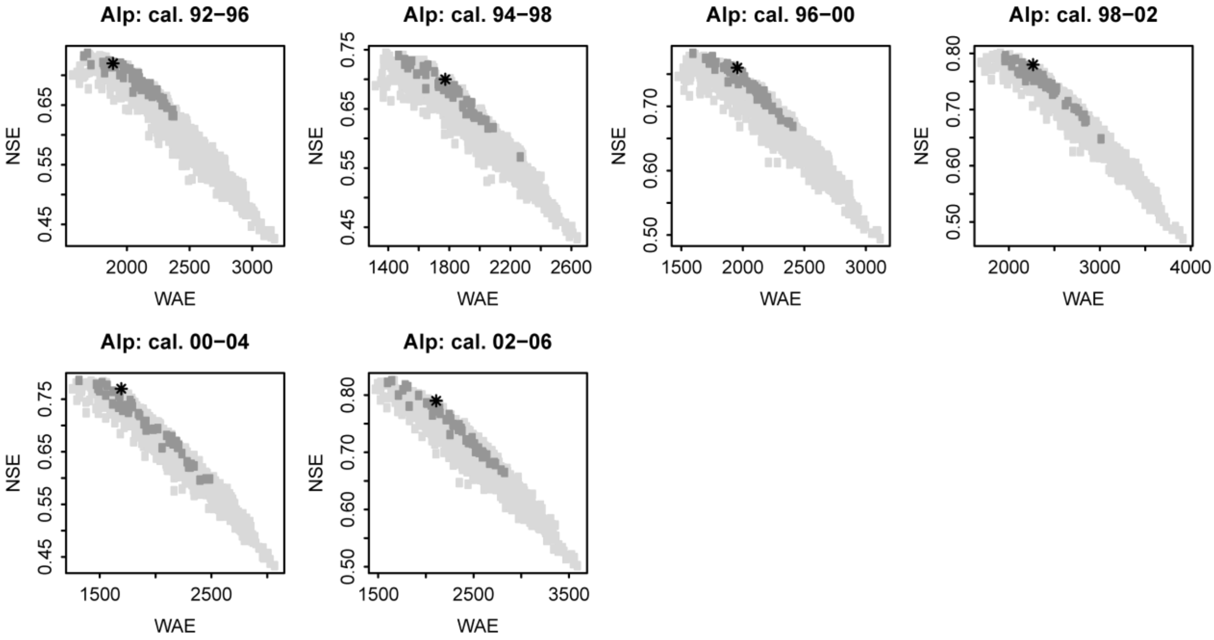

Figure A1: Verification of model calibrations for the Alp catchment. Nash-Sutcliffe efficiency (NSE) and weighted absolute errors (WAE) for simulations using the parameter sets resulting from model calibration (black star) and for simulations using 1000 randomly generated parameter sets (light grey circles), the 50 best performing plotted in dark grey.

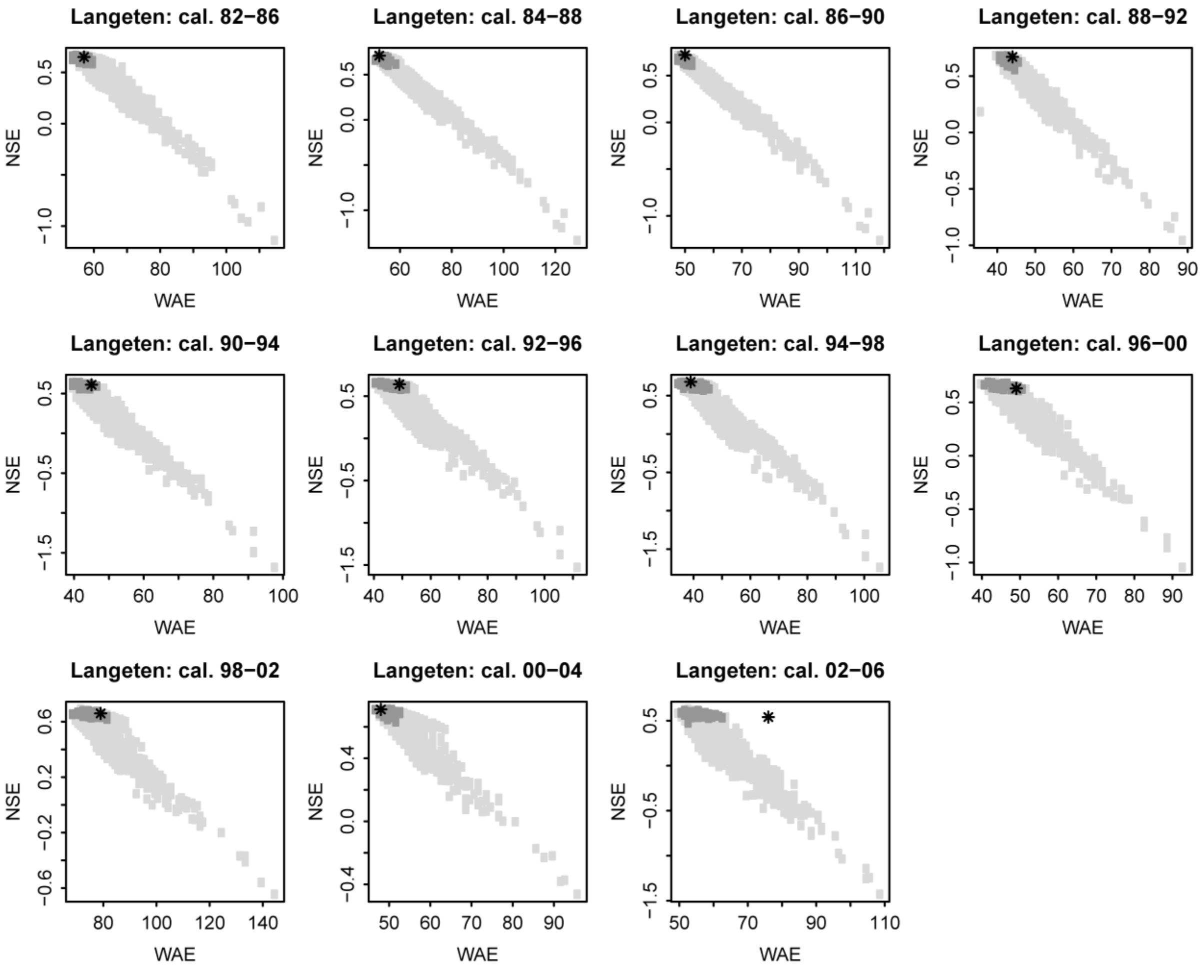

Figure A2: Verification of model calibration for the Langeten catchment. Explanation see Figure A1.

REFERENCES

Aldrich, J., 1997. R. A. Fisher and the making of maximum likelihood 1912-1922. Statistical Science, 12(3): 162-176.

BAFU, Bundesamt für Umwelt, 2008. Alp - Einsiedeln (2609), Hydrologisches Untersuchungsgebiet. BAFU Abteilung Hydrologie, Bern.

BAFU, Bundesamt für Umwelt, 2008. Langeten - Huttwil, Häberenbad (2343), Hydrologisches Untersuchungsgebiet. BAFU Abteilung Hydrologie, Bern.

Barredo, J.I., 2007. Major flood disasters in Europe: 1950-2005. Natural Hazards, 42(1): 125-148.

Borga, M., Boscolo, P., Zanon, F. and Sangati, M., 2007. Hydrometeorological analysis of the 29 August 2003 flash flood in the Eastern Italian Alps. Journal of Hydrometeorology, 8(5): 1049-1067.

Borga, M., Tonelli, F., Moore, R.J. and Andrieu, H., 2002. Long-term assessment of bias adjustment in radar rainfall estimation. Water Resource Research, 38(11): 1226.

Burch, H., 1994. Ein Rückblick auf die hydrologische Forschung der WSL im Alptal. Beiträge zur Hydrologie der Schweiz, 35: 18-33.

BWG, Schweiz Bundesamt für Wasser und Geologie, 2001. Hochwasserschutz an Fliessgewässern Wegleitung. BBL/EDMZ, Bern, 72 pp.

Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer, London, 208 pp.

Collier, C.G., 2007. Flash flood forecasting: What are the limits of predictability? Quarterly Journal of the Royal Meteorological Society, 133(622): 3-23.

Creutin, J.D. and Borga, M., 2003. Radar hydrology modifies the monitoring of flash-flood hazard. Hydrological Processes, 17(7): 1453-1456.

Delrieu, G., Braud, I., Berne, A., Borga, M., Boudevillain, B., Fabry, F., Freer, J., Gaume, E., Nakakita, E., Seed, A., Tabary, P. and Uijlenhoet, R., 2009. Weather radar and hydrology Preface. Advances in Water Resources, 32(7): 969-974.

DVWK, Deutscher Verband für Wasserwirtschaft Fachausschuss "Extreme Abflüsse" und Arbeitskreis "Niederschlag-Abfluss-Modelle", 1999. Hochwasserabflüsse. Kommissionsvertrieb Wirtschafts- und Verlagsgesellschaft Gas und Wasser, Bonn, 254 pp.

Fuentes, M., Reich, B. and Lee, G., 2008. SPATIAL-TEMPORAL MESOSCALE MODELING OF RAINFALL INTENSITY USING GAGE AND RADAR DATA. Annals of Applied Statistics, 2(4): 1148-1169.

Gaume, E., Bain, V., Bernardara, P., Newinger, O., Barbuc, M., Bateman, A., Blaskovicova, L., Bloschl, G., Borga, M., Dumitrescu, A., Daliakopoulos, I., Garcia, J., Irimescu, A., Kohnova, S., Koutroulis, A., Marchi, L., Matreata, S., Medina, V., Preciso, E., Sempere-Torres, D., Stancalie, G., Szolgay, J., Tsanis, I., Velasco, D. and Viglione, A., 2009. A compilation of data on European flash floods. Journal of Hydrology, 367(1-2): 70-78.

Germann, U., Berenguer, M., Sempere-Torres, D. and Zappa, M., 2009. REAL - Ensemble radar precipitation estimation for hydrology in a mountainous region. Quarterly Journal of the Royal Meteorological Society, 135(639): 445-456.

Germann, U., Galli, G., Boscacci, M. and Bolliger, M., 2006. Radar precipitation measurement in a mountainous region. Quarterly Journal of the Royal Meteorological Society, 132(618): 1669-1692.

Gintz, D., Hassan, M.A. and Schmidt, K.H., 1996. Frequency and magnitude of bedload transport in a mountain river. Earth Surface Processes and Landforms, 21(5): 433-445.

Griffiths, P.G., Magirl, C.S., Webb, R.H., Pytlak, E., Troch, P.A. and Lyon, S.W., 2009. Spatial distribution and frequency of precipitation during an extreme event: July 2006 mesoscale convective complexes and floods in southeastern Arizona. Water Resource Research, 45(W07419): doi:10.1029/2008WR007380.

Gurtz, J., Baltensweiler, A. and Lang, H., 1999. Spatially distributed hydrotope-based modelling of evapotranspiration and runoff in mountainous basins. Hydrological Processes, 13(17): 2751-2768.

Gurtz, J., Zappa, M., Jasper, K., Lang, H., Verbunt, M., Badoux, A. and Vitvar, T., 2003. A comparative study in modelling runoff and its components in two mountainous catchments. Hydrological Processes, 17(2): 297-311.

Hegg, C., McArdell, B.W. and Badoux, A., 2006. One hundred years of mountain hydrology in Switzerland by the WSL. Hydrological Processes, 20(2): 371-376.

Hilker, N., Badoux, A. and Hegg, C., 2009. The Swiss flood and landslide damage database 1972-2007. Natural Hazards and Earth System Sciences, 9(3): 913-925.

IHW, Institut für Hydromechanik und Wasserwirtschaft, ETH Zürich, 1998. Hochwasserschutz Einsiedeln (Untersuchungen über die Grinnekapazität und die Grösse extremer Hochwasser der Alp), Zürich.

Jaun, S. and Ahrens, B., 2009. Evaluation of a probabilistic hydrometeorological forecast system. Hydrology and Earth System Sciences, 13(7): 1031-1043.

Jenkinson, A.F., 1955. THE FREQUENCY DISTRIBUTION OF THE ANNUAL MAXIMUM (OR MINIMUM) VALUES OF METEOROLOGICAL ELEMENTS. Quarterly Journal of the Royal Meteorological Society, 81(348): 158-171.

Krajewski, W.F. and Smith, J.A., 2002. Radar hydrology: rainfall estimation. Advances in Water Resources, 25(8-12): 1387-1394.

Lamb, R., 1999. Calibration of a conceptual rainfall-runoff model for flood frequency estimation by continuous simulation. Water Resources Research, 35(10): 3103-3114.

Leadbetter, M.R., Lindgren, G. and Rootzén, H., 1983. Extremes and related properties of random sequences and processes. Springer, New York a.o., XII, 336 pp.

Martins, E.S. and Stedinger, J.R., 2000. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. Water Resources Research, 36(3): 737-744.

Merz, R. and Blöschl, G., 2008. Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information. Water Resources Research, 44(8).

MeteoSchweiz, 2007. Witterungsbericht Juni 2007, MeteoSchweiz, Zürich.

Morin, E., Jacoby, Y., Navon, S. and Bet-Halachmi, E., 2009. Towards flash-flood prediction in the dry Dead Sea region utilizing radar rainfall information. Advances in Water Resources, 32(7): 1066-1076.

Pappenberger, F. and Beven, K.J., 2006. Ignorance is bliss: Or seven reasons not to use uncertainty analysis. Water Resources Research, 42(5).

Perry, C.A., 2000. Significant floods in the United States during the 20th Century – USGS measures a century of floods. Fact Sheet 024-00.

Rickenmann, D., 2001. Comparison of bed load transport in torrents and gravel bed streams. Water Resources Research, 37(12): 3295-3305.

Rickenmann, D. and McArdell, B.W., 2007. Continuous measurement of sediment transport in the Erlenbach stream using piezoelectric bedload impact sensors. Earth Surface Processes and Landforms, 32(9): 1362-1378.

Rusjan, S., Kobold, M. and Mikoš, M., 2009. Characteristics of the extreme rainfall event and consequent flash floods in W Slovenia in September 2007. Natural Hazards and Earth System Sciences, 9(3): 947-956.

Scherrer, S. and Naef, F., 2003. A decision scheme to indicate dominant hydrological flow processes on temperate grassland. Hydrological Processes, 17(2): 391-401.

Schiemann, R., Erdin, R., Willi, M., Frei, C., Berenguer, M. and Sempere-Torres, D., 2011. Geostatistical radar-raingauge combination with nonparametric correlograms: methodological considerations and application in Switzerland. Hydrology and Earth System Sciences, 15(5): 1515-1536.

Schmocker-Fackel, P., Naef, F. and Scherrer, S., 2007. Identifying runoff processes on the plot and catchment scale. Hydrology and Earth System Sciences, 11(2): 891-906.

Smith, R.L., 1985. Maximum likelihood estimation in a class of non-regular cases. Biometrika, 72(1): 67-90.

Turowski, J.M., Yager, E.M., Badoux, A., Rickenmann, D. and Molnar, P., 2009. The impact of exceptional events on erosion, bedload transport and channel stability in a step-pool channel. Earth Surface Processes and Landforms, 34(12): 1661-1673.

USGS, 2008. Flood Definitions. U.S. Geological Survey, Kansas Water Science Center, available at: http://ks.water.usgs.gov/waterwatch/flood/definition.html.

Viviroli, D., Zappa, M., Gurtz, J. and Weingartner, R., 2009. An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools. Environmental Modelling & Software, 24(10): 1209-1222.

Viviroli, D., Zappa, M., Schwanbeck, J., Gurtz, J. and Weingartner, R., 2009. Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland - Part I: Modelling framework and calibration results. Journal of Hydrology, 377(1-2): 191-207.

Warburton, J., 1992. Observations of Bed Load Transport and Channel Bed Changes in a Proglacial Mountain Stream. Arctic and Alpine Research, 24(3): 195-203.

Wilcox, B.P., Rawls, W.J., Brakensiek, D.L. and Wight, J.R., 1990. Predicting runoff from rangeland catchments: A comparison of two models. Water Resources Research, 26: 2401-2410.

Zappa, M. and Gurtz, J., 2003. Simulation of soil moisture and evapotranspiration in a soil profile during the 1999 MAP-Riviera Campaign. Hydrology and Earth System Sciences, 7(6): 903-919.

Zappa, M., Jaun, S., Germann, U. and Walser, A., 2009. Superposition of three sources of uncertainties in operational flood forecasting chains in mountainous areas. Atmospheric Research, (Thematic Issue on COST731): submitted.

Zappa, M. and Kan, C., 2007. Extreme heat and runoff extremes in the Swiss Alps. Natural Hazards and Earth System Sciences, 7(3): 375-389.

Zappa, M., Pos, F., Strasser, U., Warmerdam, P. and Gurtz, J., 2003. Seasonal water balance of an Alpine catchment as evaluated by different methods for spatially distributed snowmelt modelling. Nordic Hydrology, 34(3): 179-202.

Zappa, M., Rotach, M.W., Arpagaus, M., Dorninger, M., Hegg, C., Montani, A., Ranzi, R., Ament, F., Germann, U., Grossi, G., Jaun, S., Rossa, A., Vogt, S., Walser, A., Wehrhan, J. and Wunram, C., 2008. MAP D-PHASE: real-time demonstration of hydrological ensemble prediction systems. Atmospheric Science Letters, 9(2): 80-87.

Zehe, E. and Sivapalan, M., 2009. Threshold behaviour in hydrological systems as (human) geo-ecosystems: manifestations, controls, implications. Hydrology and Earth System Sciences, 13(7): 1273-1297.

AII

IFKIS-Hydro Sihl: Ein operationelles Hochwasservorhersagesystem für die Stadt Zürich und das Sihltal

■ M. Zappa, S. Jaun, A. Badoux, J. Schwanbeck, N. Addor, K. Liechti, I. Roeser, A. Walser, D. Viviroli, S. Vogt, M. Gerber, J. Trösch, R. Weingartner, M. Oplatka, G.R. Bezzola, J. Rhyner

Zusammenfassung

Als Folge der Hochwasserabflüsse der Shil im August 2005 wurden hydrologische Studien durchgeführt die zeigen, dass am Standort Zürich Hochwasserereignisse mit Spitzen von 310 bis 600 m^3s^{-1} möglich sind. Aufgrund dieser Erkenntnisse beschloss das Amt für Abfall, Energie und Luft AWEL, ein regionales Hochwasserwarnsystem zur Verbesserung der Hochwassersicherheit der Stadt Zürich und des Sihltals zu realisieren. Das Kernstück des Hochwasservorhersagesystems für die Sihl ist das hydrologische Modell PREVAH. Zusammen mit dem hydraulischen Modell FLORIS bildet es die Modellkette für das auf die Stadt Zürich massgeschneiderte Warnsystem. Die für die Initialisierung des hydrologischen Modells erforderlichen meteorologischen und hydrologischen Daten werden aus einer gemeinsamen Datensammlung entnommen, die im Rahmen des Projekts IFKIS-HYDRO lanciert wurde. Um eine nützliche Vorwarnzeit zu erhalten (z.Z. bis zu 5 Tage), wird das hydrologische Modellsystem mit Wetterdaten aus verschiedenen atmosphärischen Vorhersagemodellen angetrieben. Für die hier präsentierten Simulationen (Juni 2007 bis Dezember 2009) wurden Daten des probabilistischen COSMO-LEPS bzw. des deterministischen COSMO-7 verwendet. Zur gründlichen Analyse der Güte der Modellkette wurden vergangene Abflussvorhersagen nachgerechnet und ausgewertet. Dazu wurden verschiedene Gütemasse jeweils für den Median der mit COSMO-LEPS angetriebenen Simulation mit der von COSMO-7 angetriebenen Simulation sowie mit einem Referenzlauf (interpolierte, operationell verfügbare Beobachtungen) verglichen. Die Auswertungen zeigen, dass die Referenzsimulation deutlich bessere Ergebnisse erzielt als jegliche Vorhersage. Was die Vorhersageprodukte anbelangt, schneiden die probabilistischen Abflussprognosen generell besser ab als die deterministischen Vorhersagen. Die Gütemasse sind für die COSMO-LEPS angetriebenen Simulationen besser und verschlechtern sich langsamer mit zunehmender Vorlaufzeit als die entsprechenden Werte für die mit COSMO-7 angetriebenen Simulationen. Die Verifikation der Modellkette deutet darauf hin, dass probabilistische Abflussvorhersagen als Entscheidungsgrundlage für die Massnahmeplanung im Einzugsgebiet der Sihl besser geeignet sind als deterministische Abflussvorhersagen.

1. Einleitung

In der Schweiz werden am Bundesamt für Umwelt BAFU seit über 20 Jahren operationelle Abfluss- und Wasserstandsvorhersagen für das Rheineinzugsgebiet erstellt. Aus der Analyse des Augusthochwassers 2005 (Bezzola und Hegg, 2007) resultierte, dass auch für mittlere und kleine Schweizer Einzugsgebiete ein Bedarf für Abflussvorhersagen besteht. Des Weiteren wurde die Notwendigkeit von längeren Vorwarnzeiten aufgezeigt. Längere Vorwarnzeiten führen jedoch zu grösseren Unsicherheiten in der Vorhersage (Pappenberger und Beven, 2006) und sind daher nicht unproblematisch. Da anhand von deterministischen Vorhersagen die Unsicherheiten in der hydrologischen Modellierung nicht quantifiziert werden können, werden neuerdings probabilistische Vorhersagen auf der Basis eines Ensembles möglicher zukünftiger Witterungsverläufe zur Hilfe genommen (Ehrendorfer, 1997). Gekoppelte hydrometeorologische Vorhersagesysteme gestützt auf Ensembles wurden in der Folge entwickelt und getestet (z.B. Verbunt et al., 2007; Jaun and Ahrens, 2009). Im Jahr 2007 folgte im Rahmen des Projektes MAP D-PHASE (Mesoscale Alpine Program – Demonstration of Probabilistic Hydrological and Atmospheric Simulation of Flood Events in the Alps; Zappa et al., 2008) eine umfangreiche Demonstration des aktuellen Standes der Möglichkeiten in der hydrologischen und meteorologischen Modellierung, in welche sowohl Entwickler und Anbieter als auch die Endnutzer der Vorhersageprodukte involviert waren. Seither haben die MeteoSchweiz, das BAFU, und die WSL eng zusammengearbeitet und erste massgeschneiderte Vorhersagesysteme entwickelt.

Im vorliegenden Beitrag präsentieren wir die Implementierung eines probabilistischen Hochwasservorhersagesystems für das voralpine Einzugsgebiet der Sihl. Hochwasserereignisse können anhand einer Reihe von täglich durchgeführten Modellierungen schon frühzeitig erkannt werden, womit die Sicherheit der Stadt Zürich im Allgemeinen sowie die Situation während den Bautätigkeiten am Bahnhof Löwenstrasse (kritische Bauphase 2008 bis 2011) im Speziellen verbessert wird.

Im nächsten Abschnitt beschreiben wir das Einzugsgebiet der Sihl. Nachfolgend wird in Abschnitt 3 das dafür entwickelte Vorhersagesystem vorgestellt. In Abschnitt 4 präsentieren wir die Verifikation des Vorhersagesystems für die Periode Juni 2007 bis Dezember 2009 (Addor, 2009; Addor et al., 2010). In einem zweiten Beitrag (Badoux et al., 2010, «Wasser Energie Luft», Dezember-Ausgabe 2010) gehen wir auf die durch die Bauarbeiten im Zürcher Hauptbahnhof momentan verschärfte Gefahrensituation in der Stadt Zürich ein. Es werden die angewendeten Werkzeuge zur Entscheidungshilfe bei der Hochwasservorhersage präsentiert sowie die Organisation der Projektgruppe, das Sicherheitsdispositiv und die zur Verfügung stehenden Handlungsoptionen diskutiert.

2. Das Sihleinzugsgebiet

2.1 Gebietsbeschreibung

Das Einzugsgebiet der Sihl bis Zürich umfasst eine Fläche von 336 km^2 und erstreckt sich bei einer mittleren Höhe von 1060 m ü.M. westlich sowie südwestlich des Zürichsees *(Bild 1)*. Das Gebiet hat seinen höchsten Punkt am Druesberg auf 2282 m ü.M. im Grenzbereich der Teileinzugsgebiete der oberen Sihl und der Minster. Am unteren Ende des Sihltals wird am Standort Zürich-Sihlhölzli auf 412 m ü.M. der Abfluss der Sihl gemessen und 2 km in nordöstliche Richtung mündet diese auf rund 404 m ü.M. in die Limmat. Das Gesamteinzugsgebiet ist zu rund 42% bewaldet. Wiesen und Weiden (32%) sowie alpine und subalpine Wiesen und Rasen (13%) stellen weitere häufig vertretene Landnutzungen dar. Unproduktive Flächen (Wasser, Fels) nehmen nur 6% der Einzugsgebietsfläche ein.

Teileinzugsgebiete	Fläche [km^2]	Teilfläche ab Auslass Sihlsee [km^2]	Mittlere Höhe [m ü. M.]	Betreiber
Minster, Euthal	59.2	-	1333	BAFU
Sihlsee, Wasserstand	155.5	-	1243	Etzelwerk AG
Sihl, Auslass-Sihlsee	155.5	-	1243	Etzelwerk AG
Alp, Einsiedeln	46.6	-	1151	BAFU
Biber, Biberbrugg	31.9	-	997	BAFU
Sihl, Blattwag	257.5	102.5	1171	Kt. Zürich
Sihl, Zürich	336.0	180.5	1047	BAFU

Tabelle 1. Gebietskenngrössen der Teileinzugsgebiete der Sihl für welche operationelle Abfluss-/Wasserstandmessung verfügbar sind (modifiziert nach Schwanbeck et al., 2007; vgl. auch Bild 1).

Aufgrund der starken Beeinflussung des Abflusses der Sihl durch die Wasserkraftnutzung lässt sich das gesamte Einzugsgebiet am Pegel Schlagen (Auslass Sihlsee in *Bild 1*) unmittelbar unterhalb des Sihl-Stausees in zwei durch den Sihlsee hydrologisch entkoppelte Hauptteile gliedern *(Bild 1)*: in einen südlich gelegenen Teil (Einzugsgebiet des Sihlsees, 155 km^2) sowie in einen nördlich des Sihlsees gelegenen, etwas grösseren Teil (rund 181 km^2) der sich bis nach Zürich erstreckt.

Das obere Einzugsgebiet umfasst knapp die Hälfte der gesamten Gebietsfläche, aber aufgrund der Höhenlage sowie des Einflusses der Orographie fallen hier im Vergleich zum unteren Teilgebiet im Mittel die grösseren Niederschlagsmengen an. Somit wurde ursprünglich auch der grössere Anteil des Abflusses der Sihl am Standort Zürich in diesem oberen Einzugsgebietsteil gebildet. Seit dem Beginn der Wasserkraftnutzung zeigt sich dies jedoch nicht mehr im Abflussregime der Sihl. Die gesamten Wassermengen der oberen Sihl, der Minster und einigen kleineren Bäche werden seit dem Jahr 1938 auf einer Höhe von rund 889 m ü.M. im Sihlsee zwischengespeichert und zu einem grossen Teil über einen 2900 m langen Druckstollen zur Stromproduktion in die Zentrale in Altendorf geführt und schliesslich in den Zürichsee geleitet. Im Pumpspeicherbetrieb wird gelegentlich Wasser vom Zürichsee in den Sihlsee gepumpt.

Diese starke Beeinflussung des Regimes der Sihl muss im hydrometeorologischen Vorhersagesystem berücksichtigt werden. Zu diesem Zweck wird das Gesamteinzugsgebiet zur Abflussmodellierung in Teileinzugsgebiete unterteilt *(Bild 1)*, wobei darauf geachtet wurde, dass diese möglichst mit einer Pegelmessstelle versehen sind. Je mehr einzelne Abflussmessreihen zur Verfügung stehen, desto besser lassen sich die freien Parameter des verwendeten hydrologischen Modells kalibrieren.

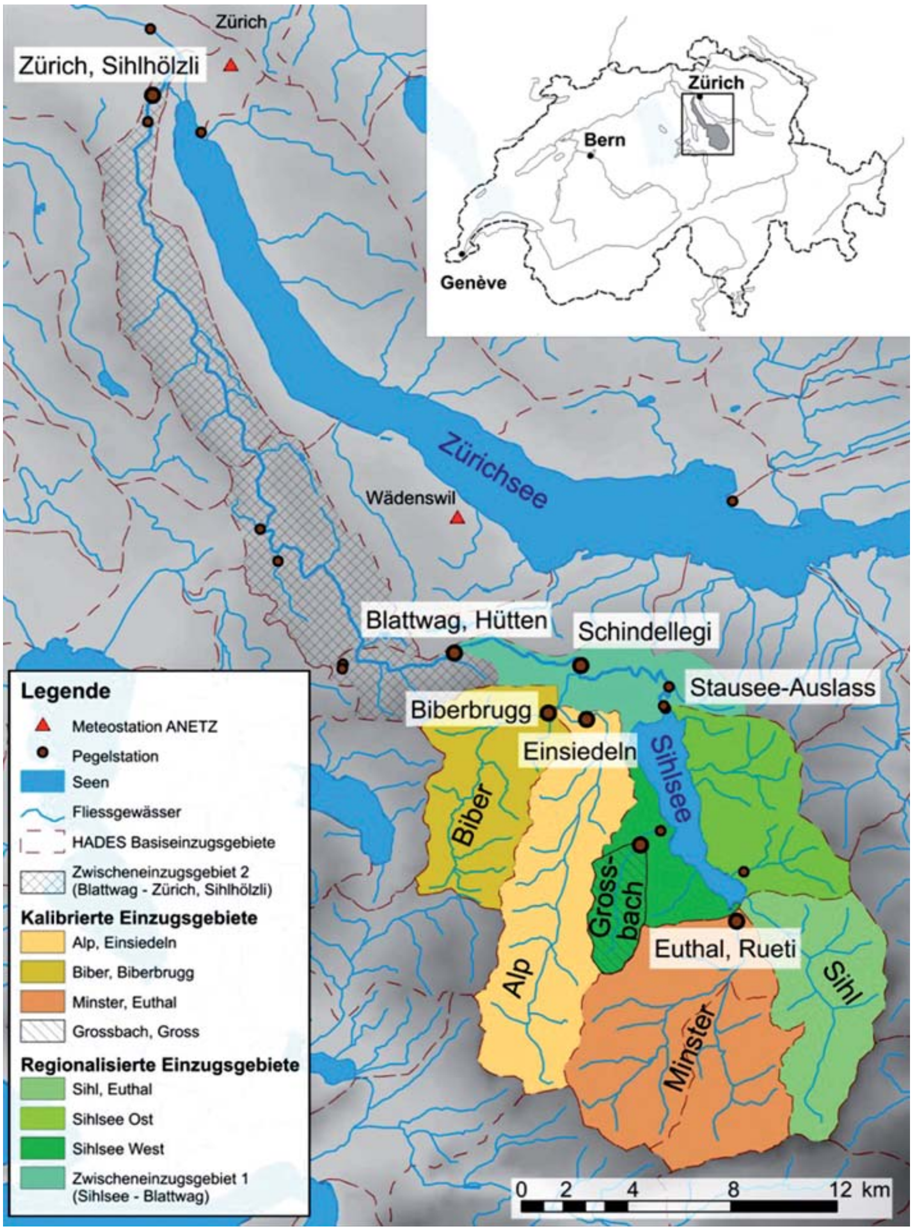

Bild 1. Übersicht über das Einzugsgebiet der Sihl und dessen Gliederung in die verschiedenen Teileinzugsgebiete (aus Schwanbeck et al., 2007; vgl. auch Tabelle 1).

2.2 Historische Hochwasser

Historische Hochwasser an der Sihl mit beträchtlichen Schäden für die Stadt Zürich und das Sihltal sind aus den Jahren

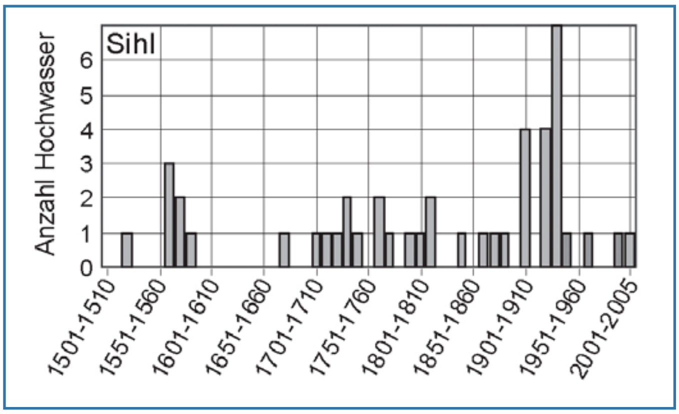

Bild 2. Historische Hochwasser der Sihl von 1501 bis 2005. Für jede Dekade wurde die Anzahl der historischen (ungemessenen) bzw. seit 1919 der gemessenen Hochwasser mit einer Abflussspitze über 200 m^3s^{-1} gezählt (aus: Schmocker-Fackel and Naef, 2010).

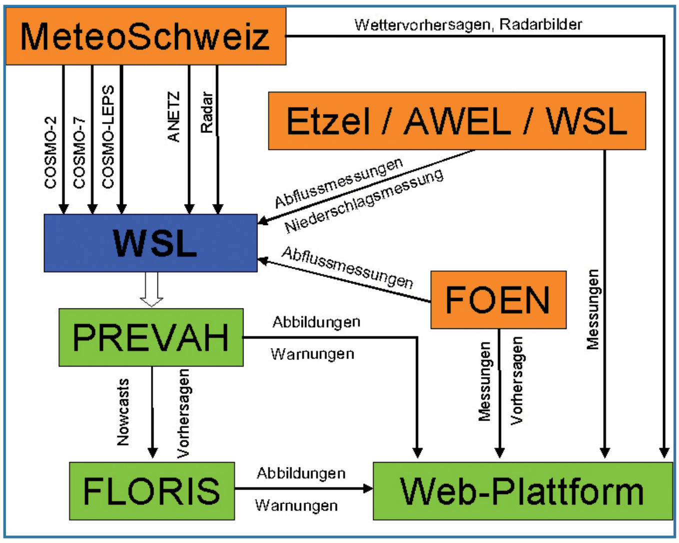

Bild 3. Schematische Darstellung der Vorhersagekette. Die orangen Felder kennzeichnen die Datenlieferanten (Dateninput); an der WSL laufen alle Daten zusammen und werden durch die Modelle PREVAH und FLORIS verarbeitet. Schliesslich werden Bilder und Auswertungen auf einer passwortgeschützten Web-Plattform veröffentlicht (siehe Badoux et al., 2010).

1846, 1874 und 1910 bekannt. Zwar gibt es aus dieser Zeit noch keine Pegelmessungen, trotzdem lässt sich die Grössenordnung dieser Ereignisse abschätzen. Dabei kann auf die Untersuchungen der Scherrer AG (2007) zurückgegriffen werden: Für das Hochwasser vom 31. Juli bis 1. August 1874 wurden Spitzenabflüsse der Sihl von 460–570 m^3s^{-1} rekonstruiert; bei dem Hochwasser im Juni 1910 sind in Zürich wahrscheinlich Höchstwerte von 380–475 m^3s^{-1} aufgetreten.

In der Studie von Schwanbeck et al. (2007) wurden mit dem kombinierten hydrologisch-hydraulischen Modellsystem PREVAH/FLORIS, das auch im hier vorgestellten operationellen Betrieb verwendet wird, extreme Niederschlagsszenarien zu den entsprechenden Abflussszenarien für die Sihl analysiert. Dabei wurde gezeigt, dass Hochwasserereignisse mit Spitzen im Bereich von 310–600 m^3s^{-1} am Standort Zürich auch unter heutigen Verhältnissen durchaus auftreten können.

In der über 90 jährigen Abflussmessreihe der Sihl traten so hohe Werte jedoch nur einmal auf. Im Jahr 1934, als der Abfluss der Sihl noch nicht durch die Wasserkraftnutzung beeinflusst war, wurden rund 340 m^3s^{-1} verzeichnet. In den darauf folgenden 71 Jahren von 1938–2008 nach Erstellung des Sihlsees stieg der Abfluss nie über die im August 2005 erreichte Marke von 280 m^3s^{-1}. Generell traten seit der Inbetriebnahme des Stausees weniger grosse Abflüsse auf als in der Zeit davor *(Bild 2)*. Während in den 1970er und 1980er Jahren alle Spitzen kleiner als 200 m^3s^{-1} waren, sind in den letzten 12 Jahren (1998–2009) drei bedeutende Ereignisse aufgetreten (1999, 2005 und 2007).

Aufgrund dieser Erkenntnisse beschloss das Amt für Abfall, Wasser, Energie und Luft AWEL, ein regionales Hochwasserwarnsystem zur Verbesserung der Hochwassersicherheit der Stadt Zürich und des Sihltals zu realisieren (vgl. Badoux et al., 2010).

3. Das PREVAH/FLORIS Hochwasservorhersagesystem für die Sihl

Das Vorhersagesystem, das für die frühzeitige Erkennung gefährlicher Hochwassersituationen entlang der Sihl entwickelt wurde, ist auf beobachtete Daten aus diversen Messnetzen (Anfangsbedingungen für Modellierungen) sowie auf Prognosedaten aus den numerischen Wettervorhersagemodellen der MeteoSchweiz angewiesen. Dabei konnte auf die wichtigen Erfahrungen während der quasi-operationellen Periode des Forschungsprojektes MAP D-PHASE zurückgegriffen werden (Zappa und Vogt, 2007). Als Kernstück wird das hydrologische Modell PREVAH eingesetzt (Viviroli et al., 2009), dem das hydraulische Modell FLORIS nachgeschaltet wird. Diese Modellkette ist eines der Module des Warnsystems für Sommernaturgefahren IFKIS-Hydro (Romang et al., 2010). Sie wurde als massgeschneidertes regionales Warnsystem für die Stadt Zürich entwickelt und implementiert.

3.1 Anwendung von hydrometeorologischer Echtzeitinformation

Meteorologische und hydrologische Messnetze

Ein wichtiges Anliegen bei der Bereitstellung der meteorologischen Datengrundlagen ist die koordinierte Nutzung der verschiedenen existierenden hydrologischen und meteorologischen Messnetze sowie die Speicherung und Darstellung der Daten in einem gemeinsamen Informationssystem. Im Rahmen des Projektes IFKIS-HYDRO (Romang et al., 2010) wurde zum ersten Mal eine koordinierte Verwertung der Daten der Messnetze des BAFU (Abflussdaten), der MeteoSchweiz und des SLF (IMIS-Stationen) erreicht. Seit neustem werden in der Schweiz auch Daten aus kantonalen Messnetzen sowie zusätzlichen Quellen, wie Wasserkraftwerken und Forschungsinstituten, in die gemeinsamen Datensammlungen aufgenommen und für die Bestimmung der Anfangsbedingungen von hydrologischen Vorhersagen verwendet.

Die Schweiz verfügt im Allgemeinen über ein relativ dichtes Netz von meteorologischen Stationen. Für den operationellen Betrieb von hydrologischen Modellen zur Ausgabe von akkuraten Abflussvorhersagen in mittleren bis kleinen Einzugsgebieten fehlen dennoch oft repräsentative Messstationen. Dies trifft vor allem im von komplexer Topographie gekennzeichneten alpinen und voralpinen Raum zu. Als Alternative zu Niederschlagsdaten aus Bodenmessnetzen bietet die MeteoSchweiz heute räumlich und zeitlich hoch aufgelöste, quantitative Niederschlagsradardaten an (Germann et al., 2006). Probleme bestehen allerdings bei der Abdeckung von inneralpinen Regionen, wo das Radarsignal stark abgeschattet wird (Germann et al., 2006).

Zur Modellierung der Sihl werden Niederschlagsdaten in stündlicher Auf-

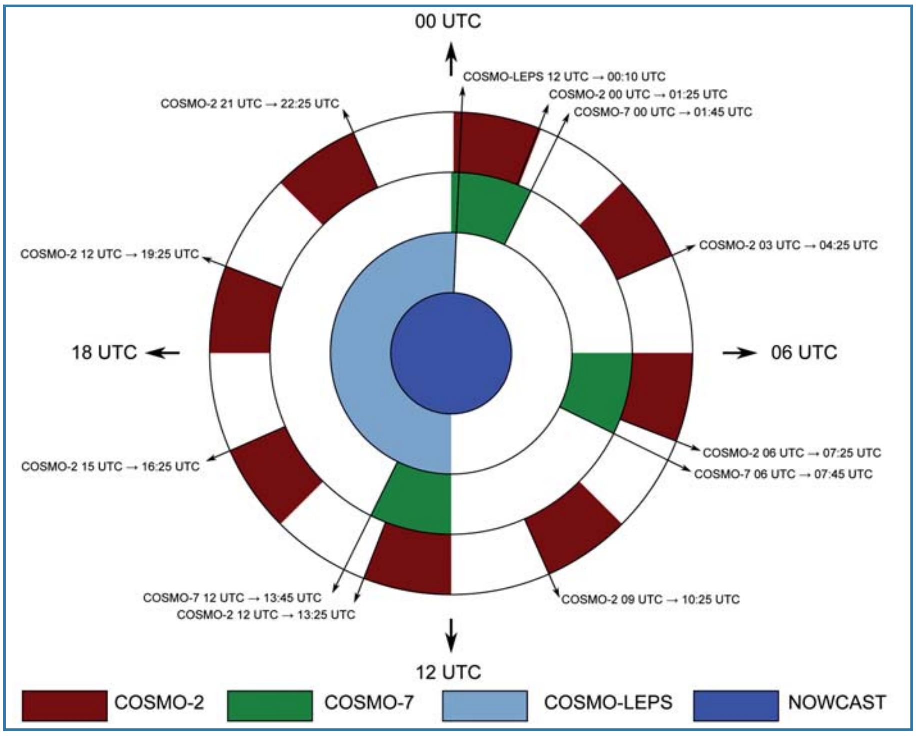

Bild 4. Modellfahrplan MeteoSchweiz (Stand April 2010, Entwurf S. Vogt)

lösung von gesamthaft zehn Stationen verwendet. Drei dieser Stationen liefern zudem alle 10 Minuten Daten, was die Ermittlung von hohen Niederschlagsintensitäten im Gebiet ermöglicht. Sieben der zehn Stationen sind mit weiteren meteorologischen Sensoren ausgerüstet und stellen auch Informationen zur Lufttemperatur, relativen Feuchte, Globalstrahlung, Windgeschwindigkeit und Sonnenscheindauer bereit. Alle diese Parameter werden, zusätzlich zum Niederschlag, dazu verwendet, das hydrologische Modell PREVAH (s. Abschnitt 3.3) anzutreiben.

Numerische Wettermodelle

Das verwendete hydrologische Modellsystem wird mit Wetterdaten aus numerischen Vohersagemodellen angetrieben. Diese modellieren den Zustand der Atmosphäre der kommenden Stunden und Tage. Derzeit betreibt die MeteoSchweiz drei Wettermodelle (COSMO-LEPS, COSMO-7 und COSMO-2), welche sich in Bezug auf die räumliche Auflösung und die Vorhersageperiode voneinander unterscheiden (Rotach, 2007).

Für die mittelfristige Frühwarnung liefert das probabilistische COSMO-LEPS täglich 16 einzelne Vorhersagevarianten (Member) der Wetterentwicklung für die kommenden fünf Tage, welche als gleich wahrscheinlich betrachtet werden. Das Modellergebnis wird als Ensemblevorhersage dargestellt, und hat sich in der Meteorologie als sehr nützlich erwiesen. Solche Vorhersagen berücksichtigen die Unsicherheiten in der Bestimmung des Anfangszustandes der Atmosphäre und lassen die Berechnung der Eintretenswahrscheinlichkeit bestimmter Ereignisse zu. Die Maschenweite von COSMO-LEPS betrug bis Dezember 2009 10 km. Seither stehen operationell COSMO-LEPS Vorhersagen mit einem horizontalen Gitterabstand von 7 km zur Verfügung.

COSMO-7 (6.6 km Maschenweite) ist ein klassisches deterministisches Wettermodell und liefert drei Mal pro Tag eine Vorhersage für die kommenden 72 Stunden. Für die kurzfristige Warnung (<24 Stunden) bietet die MeteoSchweiz alle drei Stunden die Ergebnisse des räumlich hoch aufgelösten (2.2 km Maschenweite) deterministischen Wettermodells COSMO-2 an. Dieses berechnet die Entwicklung von Konvektionszellen explizit und soll dadurch und durch die Assimilation von Niederschlagsradardaten zuverlässigere Kurzfristwarnungen vor starken Gewittern bieten.

Bild 4 stellt die zeitliche Abfolge der Rechenzeiten der für die hydrologische Vorhersage zur Verfügung stehenden numerischen Wettervorhersagen mit COSMO-2, COSMO-7 und COSMO-LEPS graphisch dar. Ein COSMO-2-Lauf trifft rund 85 Minuten nach dem Initialisierungstermin an der WSL ein *(Bild 3)*. Die benötigten umfangreichen Messungen und das Übermitteln und Aufbereiten dieser Daten benötigt allein gut 50 Minuten, während die anschliessende Berechnung der Vorhersage noch rund 30 Minuten in Anspruch nimmt. Die nachgeschaltete Abflussvorhersage mit PREVAH/FLORIS dauert weitere fünf Minuten, erst dann stehen die Produkte für den Endbenutzter bereit. COSMO-7-Vorhersagen benötigen 105 Minuten Rechenzeit an der MeteoSchweiz und weitere 8 Minuten für die hydrologische Vorhersage an der WSL. Bei den hydrologischen Vorhersagen entfällt der Hauptteil der Prozessorlast auf die hydraulischen Berechnungen. Die Berechnung der probabilistischen COSMO-LEPS-Vorhersagen mit 16 Membern ist nochmals deutlich aufwändiger. Die Ergebnisse der Wettermodellläufe am Europäischen Zentrum für mittelfristige Wettervorhersage (ECMWF, Reading, UK) stehen der WSL 12 Stunden nach deren Initialisierung zur Verfügung. PREVAH/FLORIS benötigt rund 45 Minuten auf einem 8-Prozessor Linux-Server, um die daraus resultierende 5-Tages-Ensemblevorhersage zu rechnen.

Abflussdaten

Innerhalb des Einzugsgebietes stehen Abflussdaten von sechs Messstationen in Echtzeit zur Verfügung *(Tabelle 1)*. Diese fliessen jedoch nicht unmittelbar in die Modellrechnungen ein. Einzig Informationen zum Pegelstand des Stausees werden stündlich assimiliert (s. Abschnitt 3.3).

3.2 Die hydrologisch-hydraulische Vorhersage

PREVAH

Das für die Abflussvorhersage der Sihl verwendete hydrologische Prognosemodell ist eine für den operationellen Betrieb weiterentwickelte Version des Einzugsgebietsmodells PREVAH (Precipitation-Runoff-Evapotranspiration HRU related Model; Viviroli et al., 2009a). Das Modellsystem ermöglicht die flächendifferenzierte Modellierung, bei der die räumlichen Unterschiede innerhalb eines Einzugsgebietes sowohl im Ablauf der hydrologischen Prozesse als auch der meteorologischen Eingangsvariablen berücksichtigt werden. PREVAH erfordert die Gliederung eines Gebietes nach Teilflächen mit ähnlichen hydrologischen Eigenschaften (HRUs). Zu den wesentlichen Grundbestandteilen des Modellsystems gehören verschiedene Teilmodelle zur Simulation der Schneedecke, allfälliger Gletscherflächen, der Interzeption von Niederschlag, der Feuchteausschöpfung durch Evapotranspiration, der Abflussbildung sowie der Abflusskonzentration.

Im operationellen Betrieb läuft die

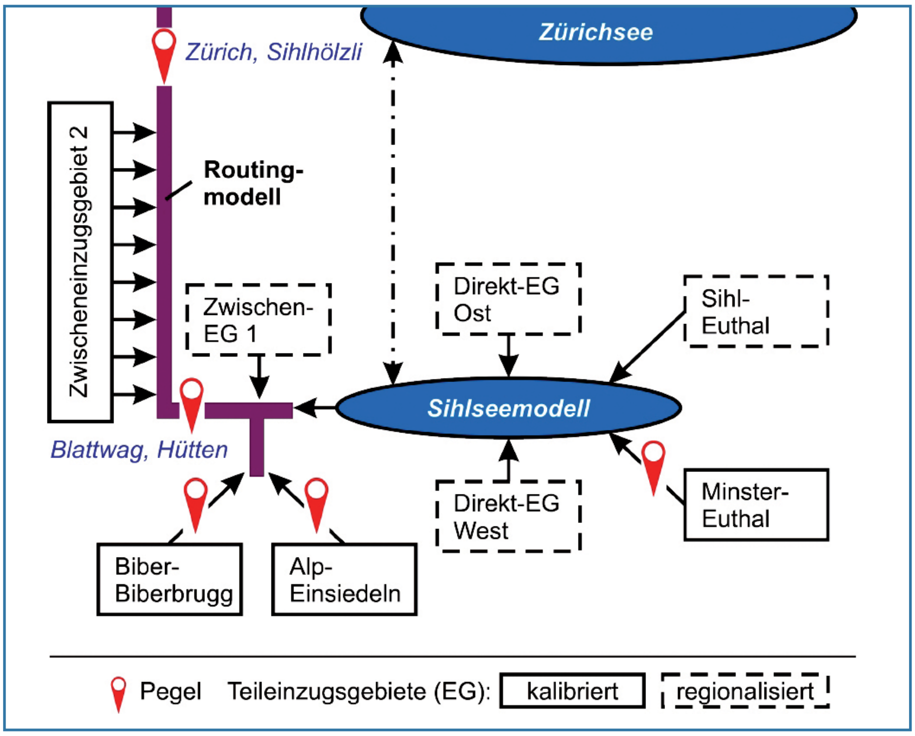

Bild 5. Aufbau des kombinierten Gesamtmodells «Sihl-Zürich» aus acht hydrologischen Teilgebietsmodellen, einem Seemodell und einem Routingmodell (nach Schwanbeck et al., 2007).

Prognoseversion von PREVAH auf einem Linux-Rechner und ist an die in *Bild 3* dargestellte Datenkette angebunden. In der Planungs- und Implementierungsphase wurden für jedes zu simulierende Teileinzugsgebiet mehrere Arbeitschritte durchgeführt: (1) Aufbereitung der Rauminformationen; (2) Definition der relevanten Wetterstationen; (3) Eichung und Verifikation des Modellansatzes mit historischen Daten; (4) Erzeugung von Dateien und Programmierung von Skripten zur Festlegung der Datenflüsse und Steuerung des Prognosemodells; (5) Anbindung an die operationelle Datenkette; (5) Nacheichung.

Seit Juni 2008 ist das PREVAH-Vorhersagesystem für die Sihl implementiert und liefert kontinuierlich Vorhersagen. Ab 2010 wird die Datenkette als eines der ersten Regionalmodule auch in das Vorhersagesystem FEWS (für «Flood Early Warning System») des BAFU implementiert. Die Simulationsergebnisse werden zudem auf der am 1. März 2010 lancierten Gemeinsamen Informationsplattform Naturgefahren des Bundes (GIN) dargestellt.

PREVAH-Kalibrierung

Die Kalibrierung des hydrologischen Modells für die neun Teileinzugsgebiete der Sihl erfolgte im Rahmen der Studie von Schwanbeck et al. (2007) zur Analyse des Hochwasserereignisses im August 2005. Für jedes Teileinzugsgebiet wurden 12 freie – das heisst justierbare – gebietsspezifische Parameter von PREVAH bestimmt. Dies erfolgte entweder durch Kalibrierung gegenüber beobachteten Abflussganglinien (Minster, Alp und Biber) oder durch Regionalisierung nach der Methodik von Viviroli et al. (2008) *(Bild 5)*. Während der Kalibrierung werden die freien Parameter durch Minimierung einer objektiven mathematischen Funktion, welche die Abweichungen zwischen Beobachtung und Simulation beschreibt, optimiert (Viviroli et al., 2009b). Bei der Eichung wurden Parameterkombinationen bevorzugt, die sowohl für Hochwasser wie auch für mittlere Abflüsse robuste Resultate liefern. Folglich zeigen simulierte Niedrigwasserabflüsse im operationellen Betrieb gelegentlich schlechtere Resultate.

Für diejenigen Teileinzugsgebiete für die keine Beobachtungsdaten vorliegen, wurden die freien Modellparameter mit dem speziell zur Anwendung von PREVAH entwickelten Regionalisierungsverfahren (Viviroli et al., 2008) bestimmt. Die Grundlage dieses Verfahrens bildet eine Datenbank von 140 mesoskaligen Schweizer Einzugsgebieten, die erfolgreich mit dieser Methode kalibriert wurden. Die in der Datenbank gespeicherten Parameter können mittels Kriging, Regressionsbeziehungen oder Ähnlichkeitsbetrachtungen für jedes beliebige mesoskalige Gebiet innerhalb des Einzugsgebietes des Rheins verwendet werden.

FLORIS

Um den genauen Verlauf von Hochwasserwellen im lang gezogenen Teileinzugsgebiet unterhalb der Messstelle Blattwag *(Bild 1)* besser simulieren zu können, wurde dem hydrologischen Modell PREVAH das hydraulische Modell FLORIS («Flood Routing System») nachgeschaltet. FLORIS ist ein kommerzielles 1D-Simulationssystem für die Nachbildung von Abflussprozessen in Gerinnen, das in den 1990er Jahren an der Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie (VAW) entwickelt wurde. In Zusammenarbeit mit privaten Ingenieurbüros (SCIETEC in Linz, Österreich und TK Consult in Zürich) wurde das Modell in den letzten Jahren weiterentwickelt. Im vorliegenden Projekt ermöglicht FLORIS ein Routing der Abflüsse aus den oberen Teileinzugsgebieten sowie die Bestimmung des Sihlsee-Pegels (s. Badoux et al., 2010). Die Anfangsbedingungen der FLORIS-Vorhersagen basieren auf einem Lauf des hydrologischen Modells mit beobachteten meteorologischen Daten und Messwerten des Seepegels.

Für die Durchführung der hydraulischen Berechnungen wurde das Gerinne der Sihl mit Hilfe von regelmässig aufgenommenen Querprofilen beschrieben (Diskretisierung). Im Gebiet der Stadt Zürich ab der Mündung in die Limmat betragen die Abstände zwischen zwei Querprofilen ungefähr 100 m. Entlang des Sihltals bis zur Staumauer des Sihlsees wachsen die Querprofilsabstände auf 600 m an. Für eine Berechnung mit FLORIS ist dies zuwenig (Numerik). Aus diesem Grund wurden zusätzliche Querprofile interpoliert, sodass mindestens alle 150 m ein Querprofil vorliegt. Im Flusslauf befinden sich Bauwerke wie zum Beispiel Schwellen, Wehre oder Brücken. Schwellen und Wehre können durch das Modell FLORIS simuliert werden, Brücken lassen sich nur teilweise berücksichtigen. Die Brückenpfeiler, die sich im Gerinne befinden und somit das Profil verkleinern, haben einen Einfluss auf den Abfluss. Hingegen kann die Unterkante einer Brücke nicht im Modell abgebildet werden. Der Sihlsee wird stark vereinfacht mit 21 Querprofilen diskretisiert, die mit Hilfe der Höhenlinien der Landeskarten erstellt wurden.

Kopplung PREVAH/FLORIS

Die vorhergesagte Ganglinie der Minster, die regionalisierten Teilgebietsganglinien der weiteren Zuflüsse sowie der Direktniederschlag auf die Seeoberfläche (abzüglich der Verdunstung) sind die entscheidenden Grössen für die Bilanzierung des

Sihlseemodells. Zusätzlich wird dem See Wasser durch Turbinierung (geschätzt basierend auf Regressionsanalysen einer zweijährigen Datenreihe) entzogen und in den Zürichsee geleitet. In die Sihl unterhalb der Staumauer wird mindestens soviel Wasser entlassen, wie nötig ist, um die gesetzlich festgelegten Restwassermengen zu garantieren. Die Bereiche oberhalb und unterhalb der Staumauer sind hydrologisch und hydraulisch erst dann vollständig gekoppelt, wenn während grossen Niederschlagsereignissen und bei einem hohen Seestand das Betriebsreglement des Stausees ab einem Seestand von 888.70 m ü.M. in Kraft tritt und in der Folge grössere Wassermengen in die Sihl abgelassen werden müssen (Badoux et al., 2010). Die Turbinierung, die erforderliche Menge an Restwasser und das Betriebsreglement wurden in einem zwischengeschalteten Modul abgebildet.

Unterhalb des Auslasses aus dem Sihlsee werden die Fliessvorgänge im Gerinne der Sihl unter Berücksichtigung der Wehre und Schwellen durch das Modell FLORIS berechnet. Alle simulierten Abflüsse aus den oberhalb liegenden sowie angrenzenden Einzugsgebieten fliessen in diese Berechnungen ein. Bis zur Abflussmessstation Blattwag-Hütten – der ersten Vergleichsmöglichkeit unterhalb des Sihlsees – sind dies der (gesteuerte) Abfluss aus dem Sihlsee, die natürlichen Abflüsse von Alp und Biber sowie die Beiträge des Zwischeneinzugsgebietes Sihlsee-Blattwag (vgl. *Bild 1*).

4. Verifikation der operationellen Modellkette

Die in *Bild 3* schematisch illustrierte Modellkette ist seit September 2008 im operationellen Betrieb. Sobald die MeteoSchweiz eine neue Vorhersage eines ihrer Wettermodelle liefert, werden die entsprechenden aktuellen Anfangsbedingungen für das hydrologisch-hydraulische Modell berechnet. Die beobachteten Klimaparameter werden mit den Klimaparametern aus den COSMO-Modellen verknüpft, um eine hydrologische Vorhersage zu berechnen.

Zur gründlichen Analyse der Güte der Modellkette wurden Vorhersagen für die Periode vom 1. Juni 2007 bis zum 31. Dezember 2009 (insgesamt 945 Vorhersagen) nachgerechnet und ausgewertet. Da die Sicherheitsverantwortlichen für die Planung von Interventionsmassnahmen auf Vorhersagen mit langer Vorlaufzeit angewiesen sind (Badoux et al., 2010), wurden zu diesem Zweck nur die Vorhersagen von COSMO-7 (00:00 UTC-Lauf) und COSMO-LEPS verwendet.

Zum besseren Verständnis der Abbildungen (*Bilder 6* bis *9*) fokussieren die folgenden Abschnitte auf die Beschreibung der Ergebnisse für die Periode von Juni bis Dezember 2007. Während dieser Zeitspanne ist am 8./9. August 2007 das grösste Ereignis der gesamten Zeitreihe aufgetreten. Zudem werden die Vorhersagen für ein weiteres Hochwasserereignis des Jahres 2008 (15./16. August) im Detail diskutiert. Eine umfassendere Darstellung der Ergebnisse findet sich in Addor (2009) und Addor et al. (2010).

4.1 Statistische Kennwerte für die gesamte Zeitreihe

Die Güte der Vorhersagen der komplexen Modellkette kann nicht als Ganzes und nicht allein durch eine einzelne Kennzahl wiedergegeben werden. Zu diesem Zweck müssen mindestens die Abflussvorhersage für den Standort Zürich und die Vorhersage des Sihlsee-Pegels beurteilt werden. Zudem ist es unerlässlich, die Güte der Prognosen für unterschiedliche Vorlaufzeiten zu evaluieren.

Station	HREF	LT1	LT2	LT3	LT4	LT5
Abfluss Zürich [m^3/s]						
E: COSMO-LEPS Med.	0.87	0.69	0.44	0.26	0.2	0.11
E: COSMO7		0.55	0.2	-0.09	-	-
DV: COSMO-LEPS Med.	1.17	1.18	1.18	1.12	1.09	1.03
DV: COSMO7		1.12	1.7	1.12	-	-
MAE: COSMO-LEPS Med.	3	3.9	5.2	5.6	6	6.3
MAE: COSMO7		4.1	5.7	6.2	-	-
Sihlsee [cm]						
E: COSMO-LEPS Med.	0.84	0.65	0.65	0.58	0.52	0.46
E: COSMO7		0.64	0.56	0.46	-	-
MAE: COSMO-LEPS Med.	1.7	2.4	5.9	9.2	12.5	15.9
MAE: COSMO7		2.4	6.3	10.1	-	-

Tabelle 2. Verifikation der simulierten Daten anhand der gemessenen Datenreihen für die Periode von Juni 2007 bis Dezember 2009. Die Tabelle gibt eine Zusammenstellung aller statistischen Kennwerte für sämtliche Vorlaufzeiten (LT1 bis LT5 für COSMO-LEPS und LT1 bis LT3 für COSMO-7) sowie für den Referenzlauf (HREF).

Die mit dem Wettermodell COSMO-LEPS angetriebenen probabilistischen Abflussvorhersagen ermöglichen Aussagen für ein bis fünf Tage in die Zukunft. Zur Übersicht wurden hier die statistischen Kennwerte nur für den Median der 16 Ensemble-Member bestimmt. Diese starke Vereinfachung ermöglicht unter Verlust der probabilistischen Information einen direkten Vergleich mit den deterministischen COSMO-7-Vorhersagen, welche einen Zeitraum von ein bis drei Tagen in die Zukunft abdecken. Für eine umfassende probabilistische Auswertung der verarbeiteten COSMO-LEPS-Daten verweisen wir auf Addor (2009).

Vorhersageabschnitte mit gleichen Vorlaufzeiten (z.B. Tag +3, 49 bis 72 Stunden ab Initialisierung der Vorhersage) wurden miteinander verbunden. Somit entstand eine kontinuierliche Zeitreihe von hydrologischen Vorhersagen mit identischen Vorlaufzeiten (LT), welche mit beobachteten Werten und einem Referenzlauf (HREF) des hydrologischen Modells verglichen werden konnten. HREF basiert auf Simulationen mit PREVAH/FLORIS, welche interpolierte, operationell verfügbare Beobachtungen der benötigen Klimaparameter als Antrieb benutzen. Folgende Gütemasse wurden berechnet:

- Die Effizienz (E) nach Nash and Sutcliffe (1970), welche den Mittelwert der Beobachtungen als Mass für das Modell annimmt ($E > 0$ bedeutet, dass das Modell eine Zeitreihe besser abschätzt als das Mittel der Beobachtungen).
- Die Abweichung der mittleren simulierten vom mittleren beobachteten Abfluss (DV). Werte über 1.0 deuten auf eine Überschätzung des mittleren Abflusses durch das Modell.
- Der mittlere absolute Fehler (MAE).

Tabelle 2 fasst diese statistischen Kennwerte für sämtliche Vorlaufzeiten (LT1 bis LT5) und für den Referenzlauf zusammen. Für die Bestimmung der statistischen Kennwerte der Seestandsmodellierung wird nicht von absoluten Werten ausgegangen. Der Wasserstand des Sihlsees wird stattdessen in eine Differenzganglinie (in cm gegenüber dem Vortag) transformiert. Dies ermöglicht eine robustere Aussage darüber, wie gut das Modell die Schwankungen des Sihlsees voraussagt.

Die Evaluation zeigt, dass die Referenzsimulation (HREF) deutlich besser als jegliche Vorhersage abschneidet. Die vorhergesagten meteorologischen Variablen sind wie erwartet nicht in der Lage, die Qualität der interpolierten Beobachtungen als Inputdaten für hydrologische

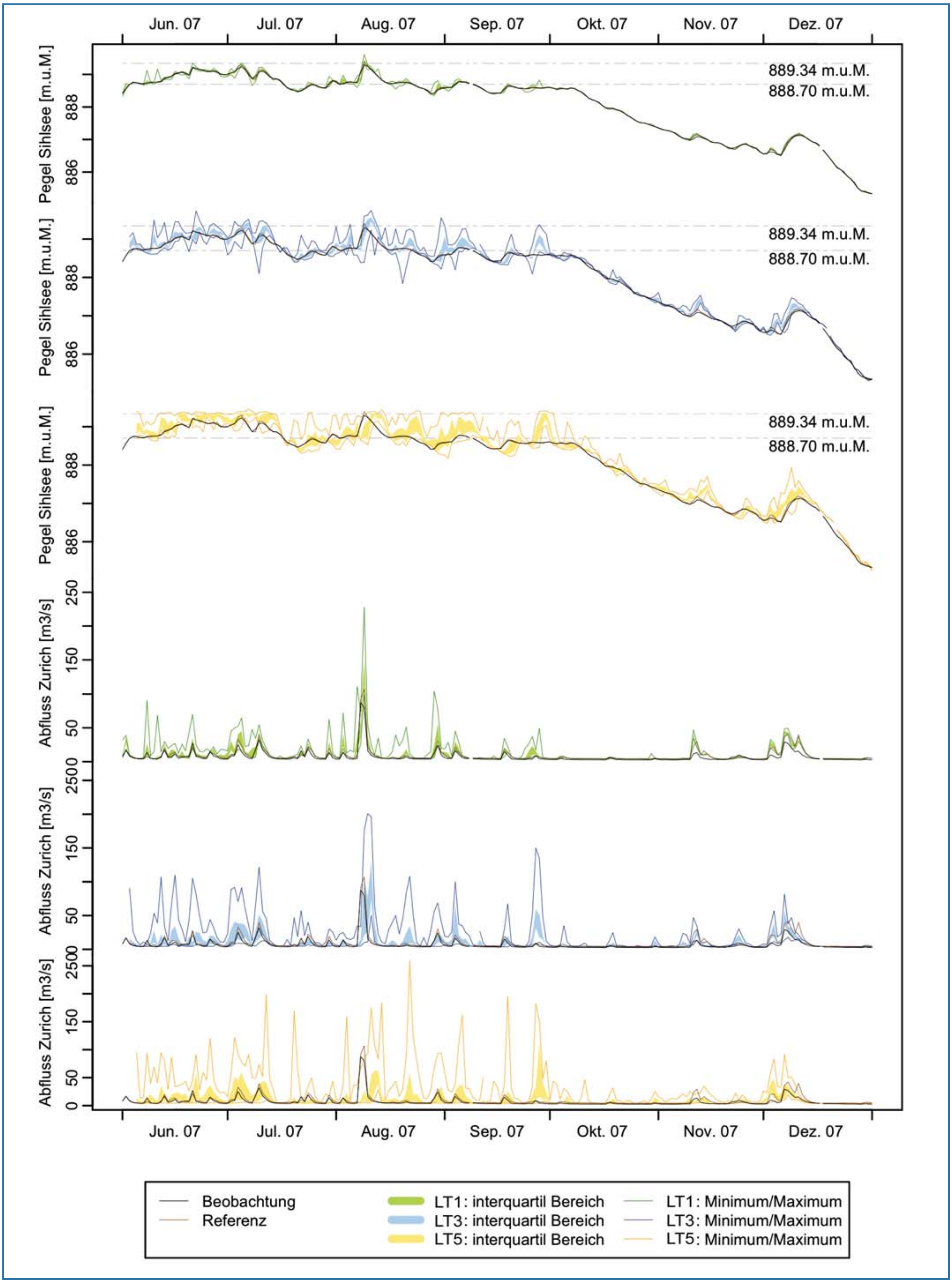

Bild 6. Angekettete COSMO-LEPS-Vorhersagen (Tagesmittel) mit unterschiedlicher Vorlaufzeit (LT, 1 Tag in grün, 3 Tage in hellblau, 5 Tage in gelb) für die Periode Juni bis Dezember 2007. Die drei oberen Graphiken beziehen sich auf die Vorhersage des Seestandes. Die Staukote (889.34 m ü.M.) und die Kote 888.7 m ü.M. (Kote Betriebsreglement bei raschem Seeanstieg) für den Sihlsee sind angegeben. Die drei unteren Gangliniengruppen beziehen sich auf Vorhersagen des Abflusses der Sihl für den Standort Zürich Sihlhölzli.

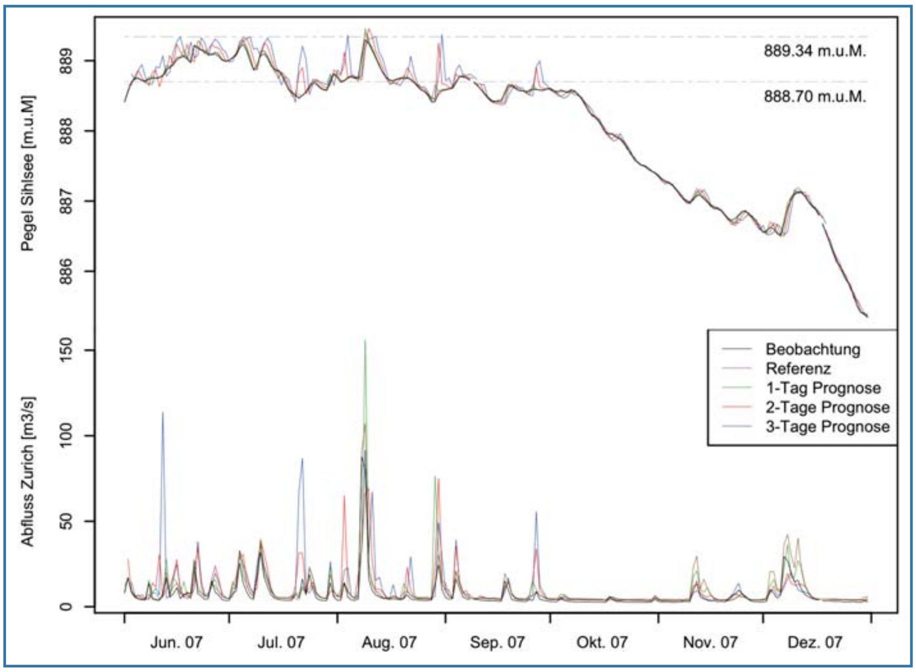

Bild 7. Angekettete COSMO-7-Vorhersagen (Tagesmittel) mit unterschiedlicher Vorlaufzeit (LT, 1 Tag in grün, 2 Tage in rot, 3 Tage in hellblau) für die Periode Juni bis Dezember 2007. Die obere Graphik bezieht sich auf die Vorhersage des Seestandes. Die Staukote (889.34 m ü. M.) und die Wehrreglementkote (888.7 m ü. M.) für den Sihlsee sind angegeben. Die untere Graphik bezieht sich auf Vorhersagen des Abflusses der Sihl für den Standort Zürich Sihlhölzli.

Simulationen zu erreichen. Was die Vorhersageprodukte betrifft, kann festgehalten werden, dass die probabilistischen Abflussprognosen (COSMO-LEPS-Ensemble und entsprechender Median) generell einen besseren Eindruck hinterlassen als die deterministischen Vorhersagen mit COSMO-7.

Die statistischen Kenngrössen E, DV und MAE sind insgesamt für die von COSMO-LEPS angetriebenen Simulationen besser und sie verschlechtern sich langsamer mit zunehmendem Vorhersagezeitraum als die entsprechenden Werte für die auf COSMO-7 basierenden Vorhersagen. Insbesondere scheinen Wahrscheinlichkeitsprognosen von COSMO-LEPS stabilere mittelfristige Prognosen bei instabilen atmosphärischen Bedingungen zu bieten. *Tabelle 2* zeigt, dass COSMO-LEPS bei Vorlaufzeiten von vier und fünf Tagen (LT4 und LT5) eine ähnliche oder bessere Güte aufweist als COSMO-7 bei einer Vorlaufzeit von 3 Tagen (LT3).

4.2 PREVAH/FLORIS COSMO Vorhersage Juni bis Dezember 2007

COSMO-LEPS

Bild 6 gibt eine qualitative Übersicht der Güte der probabilistischen Modellkette über einen längeren Zeitabschnitt von Juni bis Dezember 2007. Die Abbildung zeigt Resultate für die Station Zürich Sihlhölzli und für den Pegelstand des Sihlsees. Der Referenzlauf für die Simulation der mittleren Tageswerte zeigt eine sehr gute Übereinstimmung zwischen der PREVAH/FLORIS-Kette und den Beobachtungen. Diese visuelle Einschätzung wird durch die statistischen Kennwerte bestätigt *(Tabelle 2)*. Es ist zu beachten, dass die Simulationen des Seestandes jeden Tag bei Initialisierung der Simulation mit der Beobachtung um 00:00 Uhr abgeglichen werden. Es ist darum nicht überraschend, dass das Modell die Dynamik des Sihlsees während eines Tages gut nachbilden kann, da diese eine grosse Trägheit aufweist. Die kleinen Abweichungen zwischen HREF und den beobachteten Tagesmitteln (OBS) beruhen auf Unsicherheiten der operationellen Erfassung und der räumlichen Interpolation der beobachteten meteorologischen Variablen sowie auf Unsicherheiten im Bereich der Parametrisierung und Kalibrierung der hydrologisch-hydraulischen Modellkette.

Auch für das untere Teileinzugsgebiet bis Sihlhölzli werden mit HREF gute bis sehr gute Resultate erzielt. In grün, hellblau und gelb werden jeweils die Ergebnisse der probabilistischen Vorhersagen dargestellt. Für die Interpretation wurden Vorhersageabschnitte mit gleicher Vorlaufzeit aneinandergekettet. Dies ermöglicht eine intuitive Einschätzung der sich verändernden Vorhersage mit länger werdenden Vorlaufzeiten. Für LT1 (grün), LT3 (hellblau) und LT5 (gelb) werden unterschiedliche Bandbreiten des Interquartalbereichs berechnet. Je länger die Vorlaufszeit, desto breiter wird dieser Bereich. Dies ist ein Zeichen von wachsender Unsicherheit mit zunehmender Laufzeit in der operationellen Vorhersage. Natürlich gibt es Fälle, wie etwa im Juli 2007, wo bereits mit fünf Tagen Vorlaufszeit ein Set von Vorhersagen vorliegt, das jeweils kompakt um den berechneten Wert liegt.

Das Hochwasser am 8. und 9. August 2007 wurde ab einer Vorlaufzeit von drei Tagen zufriedenstellend vorhergesagt. Dies sowohl für den Abfluss in Zürich, wie auch für den Stand des Sihlsees, welcher knapp auf das Stauziel von 889.34 m ü.M. anstieg.

Fehlalarme sowie verpasste Ereignisse sind die beiden Hauptprobleme in der Hochwasserwarnung. Auf *Bild 6* z.B. ist Ende September ersichtlich, dass in der Prognose mit drei und fünf Tagen Vorlaufzeit ein Abflussereignis mit Seeanstieg (mit möglicher Überschreitung der Warnstufe 3) als wahrscheinlich prognostiziert wird. Für rund ein Drittel der Member wird ein Abfluss von > 100 $m^3 s^{-1}$ für den Standort Zürich angegeben. Schliesslich verzeichnete die Vorhersage am Vortag des vermuteten Ereignisses viel tiefere Werte und somit zeigte keine bedrohliche Lage an. In Wahrheit stieg der Seestand an diesem Tag kaum an und der Abfluss am Standort Zürich HB stieg nur unbedeutend.

COSMO-7

Bild 7 ist das Pendant zu *Bild 6* für mit COSMO-7-Modelldaten angetriebene Simulationen von PREVAH/FLORIS. Sowohl für den Abfluss in Zürich Sihlhölzli, wie auch für den Pegelstand des Sihlsee sind hier die deterministischen Vorhersagen mit drei, zwei und einem Tag Vorlaufzeit aneinandergekettet.

Basierend auf der visuellen Betrachtung der Ganglinien ist es schwierig zu erkennen, ob die COSMO-7-basierten hydrologischen Vorhersagen besser abschneiden als diejenigen von COSMO-LEPS. Die statistischen Kennwerte (Tabelle 2 und Abschnitt 4.1) deuten auf ein leicht besseres Resultat von COSMO-LEPS. Vergleicht man die Ganglinien für den in den *Bildern 6* und *7* dargestellten Zeitabschnitt, erkennt man keine grundlegenden Abweichungen der Modellresultate. Für den 8. und 9. August liefert auch PREVAH/FLORIS-COSMO-7 für jede der drei Vorlaufszeiten eindeutige Anzeichen dafür, dass ein bedrohliches Hochwasser bevorsteht.

Bei konvektiven Wetterlagen besteht die Tendenz, dass die deterministische Vorhersage mit zwei oder drei Tagen Vorlaufzeit bedrohliche Abflussentwicklungen anzeigt. Solche Perioden werden von PREVAH/FLORIS-COSMO-LEPS oft optimistischer eingestuft. In der Periode von Juni bis August 2007 traten solche Fälle mit prognostizierten Abflüssen von 50–100 $m^3 s^{-1}$ dreimal auf *(Bild 7)*. In allen diesen Fällen zeigten jeweils mehr als drei Viertel der Member der probabilistischen Modellkette bei jeder Vorlaufszeit unbedrohliche Abflussverhältnisse, was sich auch bewahrheitete *(Bild 6)*. Auch die mit COSMO-7 angetriebene hydrologische Modellierung überschätzt das Abflussgeschehen Ende September 2007 bei Vorlaufszeiten von mehr als einem Tag. An diesem Fall zeigt sich die Wichtigkeit einer funktionierenden Alarmorganisation, welche mit diesen Methoden der Abflussvorhersage und ihrer Interpretation vertraut ist. Zu diesem Zeitpunkt im Herbst 2007 wurden noch keine operationellen hydrologischen Modellierungen für die Sicherheit der Stadt Zürich und der Baustelle Bahnhof Löwenstrasse bereitgestellt. Andernfalls gehen die Autoren davon aus, dass die in Badoux et al. (2010) beschriebenen organisatorischen Massnahmen eingeleitet worden wären. Ein Entscheid zur Vorabsenkung des Sihlsees um 75 cm drei Tage vor dem «Nicht-Ereignis» hätte getroffen werden können. In einem solchen Fall hätte dies finanzielle Folgen (Entschädigung für die Betreiber des Etzelwerkes) mit sich gebracht.

4.3 Simulation der Dotierwassermengen Juni-Dezember 2007

Die Abflussvorhersagen können selbstverständlich auch für andere Fragestellungen als die Frühwarnung vor Hochwasser verwendet werden. Aus Sicht der Wasserwirtschaft und des Gewässerschutzes ist es zum Beispiel auch interessant zu wissen, wie gut und wie lange sich die Dotierwassermengen einhalten lassen. Die für die Sihl (Pegel Blattwag) derzeit verbindlichen Dotiermengen, die per Gesetz in den Flussabschnitt unterhalb des Stausees geleitet werden müssen, betragen je nach Saison mindestens 2.5 $m^3 s^{-1}$ (Winter), bzw. 3.0 $m^3 s^{-1}$ (Sommer). Falls die kumulierten Abflussmengen der Alp, der Biber und des restlichen Zwischeneinzugsgebiets bis Blattwag *(Bild 1)* diesen Wert nicht erreichen, müssen die Betreiber des Sihlsees die Restwassermengen vom Grundwert von rund 0.4 $m^3 s^{-1}$ erhöhen. Dieses Wasser ist ein Verlust für die Wasserkraftnutzung, aber lebenswichtig für das Ökosystem des Flusses. *Bild 8* zeigt, dass die PREVAH/FLORIS-Modellkette gut in der Lage ist, eine nötige Erhöhung des Ausflusses aus dem See zur Einhaltung der Dotiermengen mit drei Tage Vorlaufzeit vorauszusagen. Das System ist auch grösstenteils in der Lage die Abfolge von trockener und feuchter Witterung mit einer Vorlaufszeit von drei Tage widerzugeben. Die 16 simulierten Vorhersagen für die Restwassermengen (auch hier als Ensembleprognose dargestellt) zeigen mit ihrer Bandbreite die möglichen Entwicklungen der vorgeschriebenen Abgabemengen, die in drei Tagen dem Fluss zugeführt werden müssen. Obwohl momentan von geringer ökonomischer Bedeutung, könnte diese Information zukünftig im Rahmen von Optimierungsprozessen bezüglich der Wasserkraftnutzung in die Produktionsplanung einbezogen werden.

4.4 Beispiel: Ereignis vom 15./16. August 2008

Die wichtigste Herausforderung für das Modellsystem bleibt aber die zuverlässige Vorhersage von Hochwasser für die Sicherheitsverantwortlichen im Sihltal und der Stadt Zürich (Badoux et al.,2010). Da die Stadt lange vor dem Eintreffen eines Ereignisses mit der Einleitung von Notfallmassnahmen beginnen muss, werden die Vorhersagen Tag für Tag berechnet und allfällige Veränderungen (Beruhigung, Bestätigung oder Verschärfung) der jüngsten Vorhersage gegenüber der Vorhersagen der Vortage müssen den Nutzern so einfach wie möglich zugänglich gemacht werden.

Seit der Einführung der probabilistischen Vorhersagen werden zum Beispiel sogenannte Persistenztafeln zusammengestellt (Thielen et al., 2009), die zeigen, mit welchen Wahrscheinlichkeiten mit der Überschreitung eines Grenzwertes in den kommenden Tagen zu rechnen ist. *Bild 9* zeigt eine erweiterte Variante solcher Tafeln, welche für ein Ereignis an der Sihl im August 2008 nachgerechnet wurde (Addor, 2009).

Die Ordinate zeigt den Tag an dem die COSMO-LEPS-Vorhersage gestartet wurde. Die dazugehörenden Kästchen auf der Abszisse geben für jeden der fünf Tage der Vorhersage eine Information über die wahrscheinlichste Warnmeldung (innere Farbe, mit Angabe der zugehörigen Anzahl Member) und die extremste Warnmeldung (Randfarbe mit zugehörigem Maximalwert für den jeweiligen Tag). Diese Bilder lesen sich am besten «vertikal» und «von unten nach oben». Man sucht sich einen Tag in der letzten Zeile auf der X-Achse (die letzte Vorhersage) und vergleicht die Farben und Angaben mit den Vorhersagen der Vortage für denselben Zeitpunkt.

In *Bild 9* sieht man zum Beispiel, dass die «Warnstufe-Orange» (> 100 $m^3 s^{-1}$) am 15. und 16. August bereits in der Vorhersage des 12. August als möglich eingestuft wird, wobei zu diesem Zeitpunkt noch keine Maximalwerte von über 150 $m^3 s^{-1}$ erwartet wurden. Die Vorhersagen vom 13. August 2008 deuteten auf eine deutliche Verschärfung der Situation hin. Für den 15. und 16. August wurden eindeutig mehr als 60 $m^3 s^{-1}$ (gelbe Warnstufe) als wahrscheinlich erachtet. Einen weiteren Tag später (14. August 2008) erhärtete sich die Vermutung, dass der Spitzenabfluss mehr als 100 $m^3 s^{-1}$ erreichen könnte.

Bild 10 zeigt die entsprechenden «Spaghetti-Plots» (eine Linie pro COSMO-LEPS-Member) für den Sihlsee und den Standort Zürich Sihlhölzli. Zusätzlich sind die COSMO-7 Vorhersage und der Referenzlauf dargestellt. Es ist gut zu erkennen, dass bis am 16. August ein Seeanstieg um rund 80 cm und ein bedrohlicher Seestand über der Staukote von 889.34 m ü.M. als möglich erachtet wurde. Es ist aber sehr schwierig zu beurteilen, wie viele Member

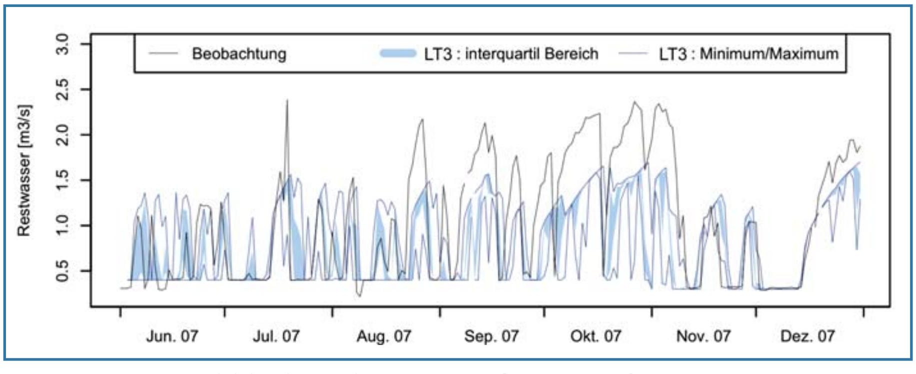

Bild 8. Angekettete COSMO-LEPS-Vorhersage (Tagesmittel) mit unterschiedlicher 3-Tage-Vorlaufzeit (TP) für die Periode Juni bis Dezember 2007. Die Graphiken beziehen sich auf die Vorhersage der Restwassermengen, welche aus dem Stausee in die Sihl (Bild 1) zur Einhaltung der Dotiermengen in Blattwag geleitet werden müssen.

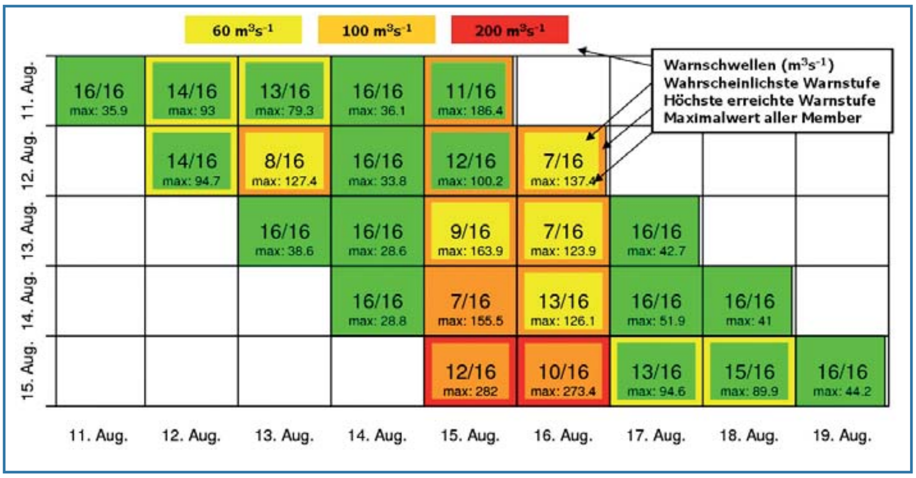

Bild 9. Persistenztafeln zur COSMO-LEPS-Vorhersage des Spitzenabflusses der Sihl in Zürich für das Ereignis des 15./16. August 2008. Die Ordinate definiert den Tag an dem die COSMO-LEPS-Vorhersage gestartet wurde. Die Abszisse definiert den Vorhersagezeitraum. In den Kästchen wird für jeden der fünf Tage der Vorhersage die entsprechende Warnstufe angegeben. Die Legende erklärt die Bedeutung der Farben (Warnstufen) und der Ziffern in den einzelnen Kästchen.

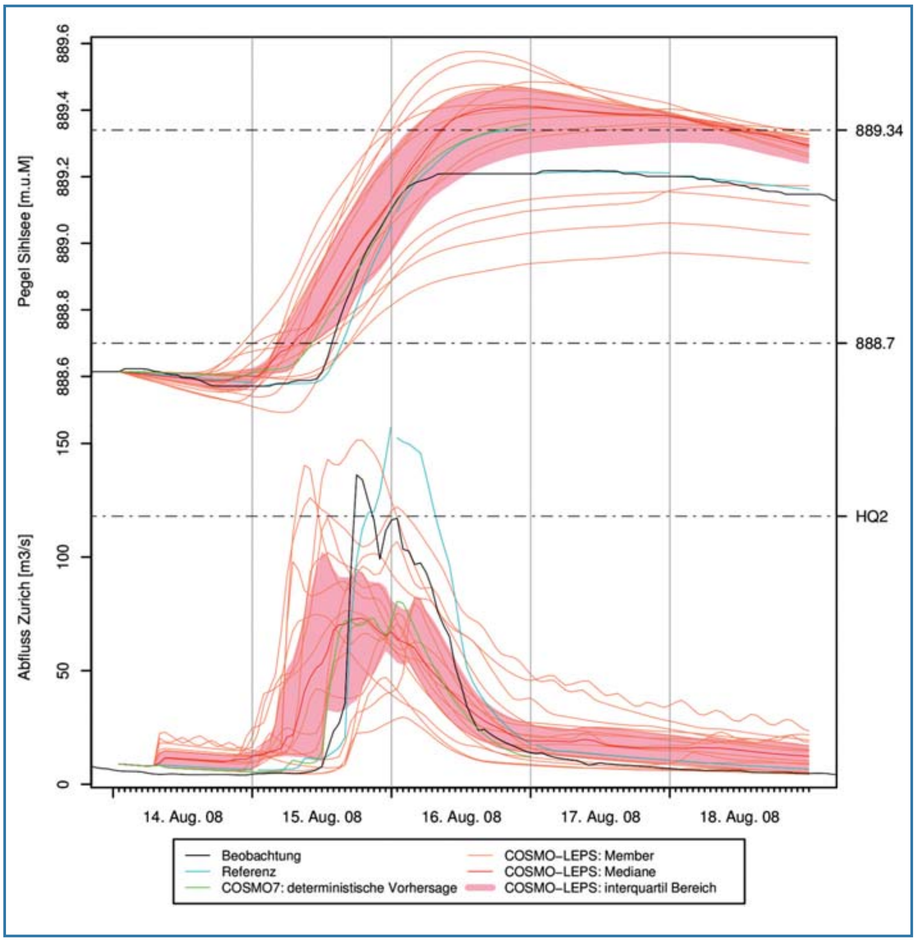

Bild 10. «Spaghetti-Plot» der a posteriori COSMO-LEPS-Vorhersage für die Periode vom 14. bis 18. August 2008. Die COSMO-7-Vorhersage und der Referenzlauf sind ebenso dargestellt. Die obere Graphik zeigt die Vorhersage des Standes des Sihlsees. Die untere Graphik die Vorhersage der Abflussganglinie für den Standort Zürich Sihlhölzli.

tatsächlich auf ein Hochwasser mit einer Wiederkehrperiode von 2 Jahre (HQ2, *Bild 10*) für Zürich Sihlhölzli deuten. Die Persistenztafel *(Bild 5)* fasst diese Informationen zusammen und zeigt, dass in 7 der 16 berechneten Szenarien mit einer Abflussspitze von über 100 $m^3 s^{-1}$ zu rechnen ist.

Die Vorhersage vom 15. August bestätigt die Vorhersage des Vortages, wonach Warnstufe «Orange» höchstwahrscheinlich sowohl am 15. wie auch am 16. August 2008 übertroffen wird. Die Abflussmessungen zeigten, dass die Schwelle von 100 $m^3 s^{-1}$ an beiden Tagen übertroffen wurden. Am 15. August wurden rund 155 $m^3 s^{-1}$ gemessen und am 16. August, bei sinkendem Pegelstand, übertraf auch eine zweite kleine Abflussspitze die HQ2 Marke von 118 $m^3 s^{-1}$.

Dieses Beispiel zeigt Möglichkeiten zur Präsentation der komplexen Information von COSMO-LEPS-Vorhersagen, die dem Endnutzer die Interpretation von probabilistischen Vorhersagen erleichtern sollen. Seit April 2010 werden die gezeigten Persistenztafeln operationell generiert und auf der IFKIS-Sihl-Informationsplattform dargestellt (Badoux et al., 2010).

5. Schlussfolgerungen

Präzise Abflussvorhersagen in alpinen und voralpinen Gebieten werden, neben der komplexen Topographie, durch die schwierige örtliche und mengenmässige Vorhersage von konvektivem Niederschlag erschwert. Während den Sommermonaten ist die Vorhersageunsicherheit auf Grund der instabilen atmosphährischen Bedingungen deshalb am höchsten. Mit Hilfe von Ensemblevorhersagen kann diese Vorhersageunsicherheit zumindest teilweise abgeschätzt werden. Die vorliegende Arbeit zeigt die erfolgreiche Integration von probabilistischer meteorologischer Information in ein operationelles hydrologisch-hydraulisches Modellsystem.

Die qualitative Evaluation mittels angekettеter Modellläufe zeigt positive Ergebnisse. Die Unsicherheitsbänder in *Bild 6* und *Bild 7* wachsen zwar mit zunehmendem Vorhersagehorizont, aber zeigen keine konstant hohe Bandbreite (Wetter-Dynamik wird abgebildet). Die Erkenntnisse aus der Verifikation der Modellkette deuten auf Vorteile von COSMO-LEPS gegenüber der deterministischen Variante COSMO-7 hin. Es lohnt sich deshalb in den höheren Rechenaufwand zu investieren, um eine bessere Entscheidungsgrundlage für die Massnahmenplanung im Einzugsgebiet der Sihl zu erhalten.

Eine Eigenheit von probabilistischen Vorhersagen ist, dass sie – anders als deterministische Vorhersagen – nach einem Ereignis nicht als «richtig» oder «falsch» eingestuft werden können, da nur die Wahrscheinlichkeit einer Überschreitung eines zu bestimmenden Schwellwertes vorhergesagt werden kann. Es muss folglich beurteilt werden, ab wann eine Vorhersage für den Endbenutzer nützlich ist. Dies kann dann auch der Fall sein, wenn die Vorhersage nach deterministischen Masstäben falsch war (Zappa und Vogt, 2007). Es ist Sache der Modellanbieter, die mit den probabilistischen Vorhersagen einhergehenden Unsicherheiten den Endnutzern zu kommunizieren. Während die deterministische Vorhersage eine eindeutige, aber möglicherweise falsche Grundlage für eine Ja-/Nein-Entscheidung liefert, bieten die probabilistischen Vorhersagen eine ehrlichere und umfassendere Einschätzung der Situation. Dies bedeutet natürlich einen einschneidenden Paradigmenwechsel in der Interpretation von Hochwasservorhersagenprodukte (vgl. *Bild 10*). Für einen nutzbringenden Einsatz der Vorhersagen ist deshalb eine kontinuierliche Schulung aller Beteiligten unerlässlich. Die kommenden Jahre werden zeigen, welche Produkte in der Praxis nützlich sind und künftig als Entscheidungshilfen für die Fachstellen, die im Krisenfall für den Hochwasserschutz zuständig sind, zur Verfügung gestellt werden sollten.

Im zweiten Beitrag («Wasser Energie Luft», Dezemberausgabe 2010) zu IFKIS-HYDRO Sihl berichten Badoux et al. (2010) über den operationellen Umgang mit dem vorgestellten Vorhersagesystem anhand des Beispiels des Baus der neuen Durchmesserlinie (Bahnhof Löwenstrasse) mitten in der Stadt Zürich.

6. Danksagung

Wir bedanken uns bei allen an diesem Projekt beteiligten Personen. Besonderer Dank geht an das Amt für Abfall, Wasser, Energie und Luft des Kantons Zürich AWEL und an das Bundesamt für Umwelt BAFU für die Finanzierung dieses innovativen regionalen Hochwasservorhersagesystems. Zudem danken wir Felix Fundel, Nadine Hilker, Mina Ossiaa und Christoph Hegg (alle WSL) sowie Uli Kuhlmann (TK Consult AG) für die Unterstützung bei der Ausführung dieses Projektes. Ingo Petzold (SLF) danken wir für die Mithilfe bei der Integration von IFKIS-Hydro Sihl in GIN.

Literatur

Addor, N. (2009): Towards flood mitigation in the Sihl catchment using operational ensemble hydrometeorological forecasts. Masterarbeit an der ETH Zurich.

Addor, N., Jaun, S., Zappa, M. (2010): Towards flood mitigation in a small mountainous catchment: evaluation of deterministic and probabilistic operational hydrometeorological forecasts. Hydrology and Earth System Sciences (submitted).

Badoux, A., Zappa, M., Schatzmann, M., Oplatka, M., Bösch, M., Jaun, S., Gross, M., Steiner, P. Hegg, C., Rhyner, J. (2010): IFKIS-Hydro Sihl: Beratung, Alarmorganisation und Handlungsmöglichkeiten während des Bau der Durchmesserlinie beim Hauptbahnhof Zürich. «Wasser Energie Luft», Ausgabe 4/2010.

Bezzola, G.R., Hegg, C. (Ed.) (2007): Ereignisanalyse Hochwasser August 2005, Teil 1 – Prozesse, Schäden und erste Einordnung. Umwelt-Wissen Nr. 0707. Bundesamt für Umwelt BAFU, Eidg. Forschungsanstalt WSL.

Ehrendorfer, M. (1997): Predicting the uncertainty of numerical weather forecasts: a review. Meteorologische Zeitschrift, 6: 147–183.

Germann, U., Galli, G.; Boscacci, M., Bolliger, M. (2006): Radar precipitation measurement in a mountainous region. Quarterly Journal of the Royal Meteorological Society, 132: 1669–1692.

Jaun, S., Ahrens, B. (2009): Evaluation of a probabilistic hydrometeorological forecast system. Hydrology and Earth System Sciences, 13: 1031–1043.

Nash, J.E., Sutcliffe, J.V. (1970): River flow forecasting through conceptual models (1), a discussion of principles. Journal of Hydrology, 10: 282–290.

Pappenberger, F., Beven, K.J. (2005): Ignorance is bliss: Or seven reasons not to use uncertainty analysis. Water Resources Research, 42: W05302, doi:10.1029/2005WR004820.

Romang, H., Zappa, M., Hilker, N., Gerber, M., Dufour, F., Frede, V., Bérod, D., Oplatka, M., Hegg, C., Rhyner, J. (2010): IFKIS-Hydro: an early warning and information system for floods and debris flows. Natural Hazards (published online), doi:10.1007/s11069-010-9507-8.

Rotach, M. (2007): Neue Entwicklungen in der Wettervorhersage: Potenzial und Anforderungen für Anwender. In: Eidgenössische Forschungsanstalt WSL (Hrsg.) Warnung bei aussergewöhnlichen Naturereignissen, Forum für Wissen 2007: 19–23.

Scherrer AG (2007): Eine Literaturauswertung der Hochwasser an der Sihl und der Limmat (1846, 1852, 1874, 1876, 1910 and 1953). Kurzbericht KA07/31 zu Handen des Amtes für Abfall, Wasser, Energie und Luft (AWEL) des Kantons Zürich, Reinach (unveröffentlicht).

Schmocker-Fackel, P., Naef, F. (2010): More frequent flooding? Changes in flood frequency in Switzerland since 1850. Journal of Hydrology, 381(1-2): 1–8.

Schwanbeck, J., Viviroli, D., Röser, I., Trösch, J., Weingartner, R. (2007): Prozessbasierte Abschätzung von Hochwassern im Einzugsgebiet der Sihl. Schlussbericht zur Studie im Auftrag des Amtes für Abfall, Wasser, Energie und Luft (AWEL) des Kantons Zürich, Bern/Zürich (unveröffentlicht).

Thielen, J., Bartholmes, J., Ramos, M.-H., de Roo, A. (2009): The European Flood Alert System - Part 1: Concept and development, Hydrology an Earth System Sciences, 13, 125–140, doi:10. 5194/hess-13-125-2009

Verbunt, M., Walser, A., Gurtz, J., Montani, A., Schär, C. (2007): Probabilistic Flood Forecasting with a Limited-Area Ensemble Prediction System: Selected Case Studies. Journal of Hydrometeorology, 8: 897–909.

Viviroli, D., Weingartner, R., Gurtz, J. (2008): Prozessbasierte Hochwasserabschätzung in ungemessenen mesoskaligen Einzugsgebieten der Schweiz «Wasser Energie Luft», 100(2): 125–130.

Viviroli, D., Zappa, M., Gurtz, J., Weingartner, R. (2009a): An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools. Environmental Modelling and Software, 24: 1209–1222.

Viviroli, D., Zappa, M., Schwanbeck, J., Gurtz, J., Weingartner, R. (2009b): Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland – Part I: Modelling framework and calibration results. Journal of Hydrology, 377(1–2): 191–207.

Zappa, M., Vogt, S. (2007): Hochwasservorhersagesysteme der neusten Generation im Praxis-Test. In: Eidgenössische Forschungsanstalt WSL (Hrsg.) Warnung bei aussergewöhnlichen Naturereignissen, Forum für Wissen 2007: 25–31.

Zappa, M., Rotach, M. W., Arpagaus, M., Dorninger, M., Hegg, C., Montani, A., Ranzi, R., Ament, F., Germann, U., Grossi, G., Jaun, S., Rossa, A., Vogt, S., Walser, A., Wehrhan, J., Wunram, C. (2008): MAP D-PHASE: Realtime demonstration of hydrological ensemble prediction systems. Atmospheric Science Letters, 9: 80–87.

Anschrift der Verfasser

M. Zappa [1], S. Jaun [1,5], A. Badoux [1], J. Schwanbeck [2,8], N. Addor [1], K. Liechti [1], I. Roeser [3,4], A. Walser [4], D. Viviroli [2,8], S. Vogt [5,4], M. Gerber [6], J. Trösch [3], R. Weingartner [2,8], M. Oplatka [7], G.R. Bezzola [5], J. Rhyner [6]

[1] Eidg. Forschungsanstalt WSL, Birmensdorf

[2] Geographisches Institut der Universität Bern GIUB, Bern

[3] TK Consult AG, Zürich

[4] MeteoSchweiz, Zürich

[5] BAFU, Bundesamt für Umwelt, Bern

[6] WSL-Institut für Schnee- und Lawinenforschung SLF, Davos

[7] AWEL, Amt für Abfall, Wasser, Energie und Luft, Zürich

[8] Oeschger-Zentrum für Klimaforschung, Universität Bern, Bern